The isolation of flowering time genes from lettuce to enable the manipulation of bolting time

Aaron Abbott

PhD Plant and Environmental Sciences

University of Warwick

Warwick HRI

September 2010

LIST OF CONTENTS

LIST OF FIGURES AND TABLES

Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

Tables

ACKNOWLEDGEMENTS

I would like to start by thanking Dr. Steve Jackson for his expert supervision; the excellent advice and guidance throughout the project has been much appreciated. It's been a thoroughly enjoyable experience.

I would also like to thank Dr. Andrea Massiah, Dr. Karl Morris and Professor Brian Thomas for their helpful inputs, advice and support. I am also extremely grateful to Jemma Taylor who has assisted me hugely throughout this project, both in the lab and in the glasshouse/field; it's been an absolute pleasure to work with you! I'd also like to thank group members Andrew Taylor, Mark Kerr, Alison Jackson, Tiziana Sgamma, Ioannis Matsoukas and Sarah Thornber.

I'd also like to thank Professor Dave Pink, Dr Paul Hand and Sandy McClement from whom I've received a huge amount of assistance and advice, especially with the EMS population work. Thanks also need to go to Peter Watson and his team in Horticultural Services at HRI for maintaining the various plant trials.

I'm also grateful to our collaborators; Professor Michelmore's group at the University of California and Rijk Zwaan, as well as David Piccaver at JEPCo with whom we ran some of our field trials. This is a BBSRC funded project, so thanks is also extended to them.

I'd like to thank my mum, dad and brothers without whose positivity, belief and unbelievable kindness I would not have achieved half of what I've managed.

Finally I'd like to thank Jane, who started the PhD as my girl and ended it as my wife! Thank you for your continual love, support and encouragement.

DECLARATION

The work referred to in this thesis is my own, unless otherwise stated and has not been submitted for a degree at another university.

SUMMARY

The time of bolting is an important factor in lettuce production because it affects the yield and quality of the harvested crop. Bolting is promoted by higher temperatures and is an increasing problem for growers with the current trend for warmer summers. Lettuce plants that are in the early stages of bolting are visibly indistinguishable from non-bolting plants, however there are changes in the biosynthesis of secondary metabolites which are produced to protect the young floral bud from insect attack. These compounds give the lettuce plant a bitter taste and render the crop unsaleable. The development of late bolting varieties, which would have a greater 'holding ability' in the field, would result in reduced crop losses and an extension to the growing season.

In many plants, the timing of the transition from vegetative growth to flowering is controlled by environmental cues which serve to communicate growth conditions favourable for sexual reproduction and seed maturation. Studies in *Arabidopsis* have led to the identification of several different pathways that come together to regulate flowering time. Little research has been done on these response pathways in lettuce, however, research has shown that components of these pathways are conserved between *Arabidopsis* and other crop species.

The aim of this project is to isolate genes regulating flowering time in lettuce in order that novel alleles of these genes can be used to manipulate bolting time. A lettuce BAC library has been screened and homologues of eight *Arabidopsis* flowering time genes, principally from the autonomous pathway, have been isolated. Functional orthologues of *FLOWERING LOCUS T* (*FT*) and the autonomous pathway gene, *FLK* have been characterised in lettuce, suggesting that there is conservation of the genes involved in flowering time in *Arabidopsis* and lettuce.

Lettuce lines with a range of bolting times, including lines which bolt significantly later than wild-type have been identified from EMS mutagenised populations of cultivated lettuce and a diversity set of wild lettuce. Homozygous lines from a *Lactuca sativa* cv. Larissa EMS population with a reproducible late bolting phenotype when tested under commercial growing conditions have been identified. These lines have been made available to Rijk Zwaan® for inclusion in future breeding programs aimed at delaying bolting and improving the 'holding' ability' of commercial lettuce crops.

Genomic sequence of selected lettuce flowering time genes have been compared between the late bolting lines and wild-type looking for polymorphisms that may account for the late bolting phenotype. Polymorphisms within these genes were identified in some of the late bolting lines, however through analysing the polymorphism in segregating backcross populations they have been shown not to be causing the late bolting phenotype. Transcriptome sequencing has also been performed to identify polymorphisms in other, possibly novel, genes which may be causing the late bolting phenotype, as yet, no mutation segregating within the late bolting lines has been identified.

ABBREVIATIONS

α	confidence level (statistics)
AGI	Arizona Genomics Institute
AGL28	*AGAMOUS-LIKE 28*
ANOVA	Analysis of Variance
AP1	*APETALA1*
APRR	*ARABIDOPSIS PSEUDO-RESPONSE REGULATOR*
At	*Arabidopsis thaliana*
B locus	bolting gene locus
BFT	*BROTHER OF FT*
BAC	Bacterial Artificial Chromosome
BC	back-cross(ed/ing)
BLAST	Basic Local Alignment Search Tool
CaMV	*Cauliflower mosaic virus*
CCA1	*CIRCADIAN CLOCK ASSOCIATED 1*
CCT	*CO*, *CO*-like and *TOC1*domains
CDF1	*CYCLING DOF FACTOR 1*
cDNA	copy DNA
CGP	Compositae Genome Project
CGN	Centre for Genetic Resources, Netherlands
CIP8	*COP INTERACTING PROTEIN 8*
CO	*CONSTANS*
COL	*CONSTANS*-like
COP1	*CONSTITUTIVE PHOTOMORPHOGENIC 1*
Col-0	Columbia
CTAB	Cetyltrimethylammonium bromide
CRY1	*CYPTOCHROME 1*
CRY2	*CYPTOCHROME 2*
Defra	Department for Environment, Food and Rural Affairs
d.f.	degrees of freedom
DCL3	*DICER-LIKE 3*
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphophates
EDTA	ethylenediaminetetraacetic acid
EB	early bolter/ing
EF1 α	*ELONGATION FACTOR 1 ALPHA*
ELF3	*EARLY FLOWERING 3*
EMS	Ethyl-Methane Sulphonate
EST	Expressed Sequence Tag

FAO	Food and Agriculture Organisation
FIO1	*FIONA1*
FKF1	*FLAVIN-BINDING, KELCH REPEAT F-BOX*
FLC	*FLOWERING LOCUS C*
FLD	*FLOWERING LOCUS D*
FLK	*FLOWERING LOCUS KH DOMAIN*
FRI	*FRIGIDA*
FT	*FLOWERING LOCUS T*
GA	gibberellic acid
Gb	giga base pairs
GI	*GIGANTEA*
GRP7	*GLYCINE-RICH BINDING PROTEIN 7*
GM	genetically modified
HD1	*HEADING DATE 1*
HD3A	*HEADING DATE 3A*
IAA	3-indolylacetic acid
iPCR	inverse PCR
JIC	John Innes Centre
K-H	K-homology
LATE1	*PEA LATE BLOOMER 1*
LB	Luria Broth
LB	late bolter/ing
LD	long day(s)
LD	*LUMINIDEPENDENS*
LDL1	*LYSINE-SPECIFIC HISTONE DEMETHYLASE*
LDL2	*LYSINE-SPECIFIC HISTONE DEMETHYLASE1*
L*er*	Landsberg *erecta*
LFY	*LEAFY*
LHY	*LATE ELONGATED HYPOCOTYL*
LOV	light, oxygen and voltage
l.s.d.	least significant difference
LsCO	*Lactuca sativa CONSTANS*
LsCOL	*Lactuca sativa CONSTANS LIKE 4*
LsCRY2	*Lactuca sativa CRYPTOCHROME 2*
LsFCA	*Lactuca sativa FCA*
LsFKF1	*Lactuca sativa FLAVIN-BINDING, KELCH REPEAT F-BOX*
LsFLC	*Lactuca sativa FLOWERING LOCUS C*
LsFLD	*Lactuca sativa FLOWERING LOCUS D*
LsFLK	*Lactuca sativa FLK*
LsFPA	*Lactuca sativa FPA*

LsFT	*Lactuca sativa FLOWERING LOCUS T*
LsFVE	*Lactuca sativa FVE*
LsFY	*Lactuca sativa FY*
LsLD	*Lactuca sativa LUMINIDEPENDENS*
MFT	*MOTHER OF FT*
miRNA	microRNA
MSI	*MULTICOPY SUPPRESSOR OF IRA1*
NaCl	sodium chloride
NaOH	sodium hydroxide
NASC	Nottingham *Arabidopsis* Seed Centre
NCBI	National Centre for Bioinformatic Information
NJ	neighbour-joining
NPK	nitrogen, phosphorus, potassium
p	probability
PCR	Polymerase Chain Reaction
PEBP	phosphatidylethanolamine-binding protein
PHD	PLANT HOMEODOMAIN
PHYA	*PHYTOCHROME A*
PHYB	*PHYTOCHROME B*
PHYC	*PHYTOCHROME C*
PHYD	*PHYTOCHROME D*
PHYE	*PHYTOCHROME E*
OsFTL	*Oryza sativa FLOWERING LOCUS T-like*
Q-Q Plots	Quantile plots
QTL	Quantitative Trait Loci
RILs	Recombinant Inbred Lines
RFT1	*RICE FLOWERING LOCUS T 1*
RKIP	Raf kinase inhibitor protein
RNA	Ribonucleic Acid
RNAi	RNA interference
RRMs	RNA Recognition Motifs
SD	short day(s)
SEM	standard error of the mean
SDS	sodium dodecyl sulphate
siRNA	small interfering RNA
SOC	*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*
SSC	saline-sodium citrate
SWIRM	SWI3p, Rsc8p, Moira
TILLING	Targeting Induced Local Lesions In Genomes
TFL1	*TERMINAL FLOWER 1*

TOC1	*TIMING OF CAB EXPRESSION 1*
Tris/ Tris-HCl	2-Amino-2-(hydroxymethyl)-1,3-propanediol
TSF	*TWIN SISTER OF FT*
WT	Wild-Type
UC Davis	University of California at Davis
UTR	untranslated region
VIN3	*VERNALISATION INSENSITIVE 3 (VIN3)*,
VRN1	*VERNALISATION 1*
VRN2	*VERNALISATION 2*
ZIML1	*ZIM-LIKE 1*
ZT	zeitgeber time
ZTL	*ZEITLUPE*

CHAPTER 1

Introduction

1.1 Project background

Food production will have to increase by 70 % over the next 40 years to meet the demand of the world's increasing population (FAO, 2009). The availability of land for crop production is on the decline, and water for agriculture is becoming increasingly scarce. Together these and other factors, including climate change will challenge the capacity of the world's food production system. Therefore the world's natural resources need to be maximised to ensure enough food is produced; this will involve increasing crop sustainability by reducing crop losses and wastage. Cultivated lettuce (*Lactuca sativa*) is an important leafy salad crop which is grown worldwide. A major source of lettuce crop losses is caused by the plant bolting before harvest. The initiation of bolting corresponds to the plant entering the reproductive phase of growth. This initiation results in the lettuce plant producing a number of secondary compounds to protect the developing floral parts from insect attack, these compounds give the plant an unpalatable bitter taste resulting in crop losses. The ability to reduce losses by delaying bolting and increasing the holding ability of lettuce plants in the field is of key interest to lettuce industry representatives, and resulted in this BBSRC funded studentship; to identify flowering time genes in lettuce to enable the manipulation of bolting.

1.2 Project aim and objectives

1.2.1 Aim

The overall aim was to identify flowering time genes that could be used to improve the holding ability of lettuce plants in the field. This would be done by identifying novel alleles of those genes which delay or prevent bolting/flowering within late bolting lines that have been isolated from a number of different lettuce populations.

1.2.2 Objectives

- To identify and isolate lettuce orthologues of targeted *Arabidopsis thaliana* flowering time genes.
- To identify late bolting (LB) commercial lettuce lines from populations containing induced mutations created by ethyl methane sulphonate (EMS) mutagenesis and retrotransposon insertions.
- To screen homozygous LB mutant lines for mutations within the lettuce flowering genes identified; if variation is not apparent within these genes then identification of mutations in non-targeted and novel flowering time lettuce genes will be investigated.
- To search for naturally occurring allelic variation in lettuce target genes that affect bolting time in a diversity set of wild lettuce varieties.
- To assess the utility of these new alleles for increased holding ability with lettuce breeders and growers.

1.3 Introduction to lettuce

1.3.1 Background and origin of crop

Lettuce belongs to the largest dicotyledonous family in the plant kingdom, the *Asteraceae* (*Compositae*). The *Compositae* is the largest and most diverse of angiosperm families, comprising of one-tenth of all flowering species (Funk *et al*., 2005). Plants within the family are characterised by compound inflorescences which consist of many tiny flowers. It contains over 40 economically important species which are grown for food, oil and medicines. Lettuce and sunflower make up two of the most economically important members of the family (Michelmore, 2007). Lettuce is a member of the subfamily *Cichorioideae*, the tribe *Lactuoideae* and the genus *Lactuca*.

Lettuce cultivars exist in a number of different varieties, the classification of which relies on differences in leaf shape and size, the degree of rosette and the way in which the head forms. Seven types of cultivated lettuce are recognised; crisphead, butterhead, cos, leaf, stem, Latin (Rodenburg, 1960) and oil-seed lettuce (Rulkens, 1987). *L. serriola* or prickly lettuce is a common wild lettuce which can be found growing world-wide. Two more well established species exist within the *Lactuca* genus; *L. saligna*, willowleaf lettuce and *L. virosa*, bitter lettuce, both of which can be found growing in Western Europe. *L. serriola* and *L. saligna* accessions exhibit either an annual or biennial life cycle, whereas *L. virosa* accessions show a biennial cycle, with a strong vernalisation requirement.

There is no evidence for the existence of a truly wild *L. sativa* (Linqvist, 1960a; Kesseli *et al*, 1991). Lindqvist, (1960a) speculated that primitive forms of cultiavated *L. sativa* descended directly from *L. serriola*. He based this assumption on the fact that the two species are so closely related; several groups of traits are

almost identical between the species as well as the chromosomes being morphologically very similar; they are basically homologous (Lindqvist, 1960b). It is likely that it was cultivated for forage or oil (Ryder, 1999). More recent studies have confirmed that cultivated lettuce is most likely to have originated by mutations which lent themselves to domestication from *L. serriola* (De Vries, 1997; Pink and Keane, 1993; Ryder, 1999). It has also been shown, through restriction fragment length polymorphism work that *L. serriola* is the most closely related species to cultivated lettuce (Kesseli *et al*., 1991).

The domestication of *L. sativa* into a form which was considered edible for humans probably took place in the Eastern Mediterranean area and/or Egypt (Ryder, 1999). Tomb paintings in Egypt which date back to around 2500BC appear to show plants which are considered by scholars to be lettuce (Harlan, 1986). The plants appear to be very similar to stem lettuce, with long, thick stems and narrow pointed leaves. Lettuce spread radiply through the Mediterranean basin, particularly in the ancient Roman and Greek civilisations. It subsequently spread to Western Europe (Ryder and Whitaker, 1976). Cos lettuce types predominated the Mediterranean area; the geographical area which is closest to the centre of origin. Cos varieties most closely resemble stem lettuce and are therefore likely to have developed from the stem types. A large amount of variability exists within cos lettuces; leaf size, shape, texture, head/non-head forming and colour. It is thought that the cos lettuce may have been the predecessor of the Batava-type crisphead (Ryder, 1999). In turn, the crisphead and iceberg varieties grown today were bred from the old Batavia types (Bohn and Whitaker, 1951). Domestication of the wild *Lactuca* species has led to a change in character combinations to satisfy human desires, for example a decrease in latex content and bitter taste, a loss of prickles and the removal of early bolting (De Vries,

1997; Ryder and Whitaker, 1976). The trend in popularity for the different lettuce types has changed over time. For example, in Northern Europe, the butterhead varieties were popular, but in the late 1970s the types which were produced began to change and there was a large increase in crisphead iceberg type lettuces. Recently consumer trends have changed again and the availability of pre-packaged leafs has seen an increase in the production a wide range of lettuce types to provide colour and texture to the product.

1.3.2 Lettuce production

Lettuce is the world's most important leafy salad vegetable (McGuire *et al*., 1993). Lettuce can be produced commercially wherever the conditions suit, it is a cool-season crop which requires good soil and an adequate water supply. Lettuce grows best in daytime temperatures of 18 °C-25 °C and a night temperature of 10 °C-15 °C. Lettuce is grown commercially throughout the world, particularly in North America, Western Europe, the Mediterranean, Australia and parts of Asia. The timing of production varies depending on the hemisphere in which production is taking place, the climate and season at any given time. Lettuce is planted outdoors in two main ways; direct drilling of seed or transplanting seedlings. Most lettuce is grown with two rows per bed 40cms apart with 1m between each bed centre. The continuous staggered sowing of lettuce varieties and types in the field is an important logistical consideration; this ensures a supply of lettuce crop throughout the growing season. Growers will also grow different varieties of lettuce suited to the time of year to ensure continuous supplies of the crop.

Lettuce is also grown under protective cover in glasshouses or plastic particularly during late autumn, winter and early spring. The protective cover allows the

conditions to be modified to suit the crop. Leaf and butterhead are the most common varieties of lettuce to be grown this way. Hydroponic production uses the nutrient film technique, with plants grown in peat blocks in concrete or plastic troughs. Nutrient solution is introduced at one end of the trough and flows by gravity through the troughs in a thin film. The plants are held in perforated covers with their roots in the solution film.

Lettuce is harvested when plants reach maturity, most growers will harvest the plants in one cutting, this reduces the amount of time taken to harvest and the amount of wastage created, and it also means that the process is less costly and more efficient. It is important to growers that plants reach maturity at the same time, meaning that the crop can be harvested in one cutting. Where large amounts of lettuce are grown it is useful to the grower that lettuce varieties have the ability to hold in the field without bolting, this holding period only needs be 2-3 days, so that field machinery and processing personnel have time to harvest the entire crop (D. Piccaver, J.E. Piccaver & Co., personal communication).

There are four main targets for lettuce breeding; they are horticultural improvement, resistance to diseases, insects and stress problems, uniformity of maturity and adaptation to specific environments (Ryder, 1999; Michelmore, 2007). The ability to increase the holding ability of lettuce plants in the field will help to reduce losses and wastage and increase productivity and sustainability.

1.3.3 Uses of Lettuce

Lettuce leaves are used principally as a raw product in salads. Lettuce is harvested and marketed in two main ways, firstly as whole lettuce plants cut and used as desired by the consumer. Secondly, as a more convenient ready cut, washed and pre-

packaged product. These products have recently increased in popularity and often comprise of a wide variety of mixtures of lettuces, endive, chicory, spinach and other vegetables. They also include other items such as salad dressing and other foodstuffs.

The large seeds from primitive forms of *L. sativa*, grown in Egypt and the Mediterranean area are pressed to express oil which is used for cooking. This ancient practice may predate the use of lettuce as a vegetable.

Stem lettuce is eaten raw in countries such as Egypt and is used as cooked vegetable in China. Latex extracts of several *Lactuca* species contain two sesquiterpene lactones called lactucin and lactucopicrin. These compounds have sedative effects. Extracts from *L. virosa* are used as sleep inducers and cough suppressants in Europe (Ryder, 1999).

Lettuce contributes vitamins and minerals to the human diet along with fibre and water. The nutritional value of lettuce varies depending on leaf colour. The more green the leaf the more nutritious it is. Cos and leaf lettuce generally exceed the butterhead and crisphead varieties in content of vitamins, minerals and fibre.

1.3.4 Economic importance of lettuce

In 2008, a worldwide area of 1.06 million hectares of lettuce and chicory was harvested, with 23.5 million metric tonnes of lettuce being produced (FAOSTAT, 2009). Provisional figures predict that 5349 hectares of lettuce will be grown in the UK in 2009-10. 117,300 tonnes of lettuce were grown with a home production value (HPM) of £84.7 million in the UK in 2009 (Defra Horticultural Statistics, 2009). Table 1.1 summarises the area of lettuce harvested, the amount of lettuce produced

and the value of the crop in the UK between 2005 and 2009. Table 1.2 describes the amount of lettuce imported to and exported from the UK between 2005 and 2009.

Year	Area harvested (Ha)	Production (thousand tonnes)	HPM (£ 1000)
2004-5	6010	131.7	71,844
2005-6	5601	126.4	96,462
2006-7	6069	109.0	81,815
2007-8	6075	116.8	97,727
2008-9	5592	117.3	84,688

Table 1.1 – Area harvested, amount of lettuce produced and the home production value recorded in the UK between 2005 and 2009. (Defra Horticultural Statistics, 2009)

Year	UK Production	Imports	Exports	Total Supply	UK Production as % of total supply
2005	131.7	188.4	6.5	313.6	42
2006	126.4	173.9	5.8	294.5	42.9
2007	109.0	178.7	3.6	284.1	38.4
2008	116.8	172.5	4.1	285.2	41
2009	117.3	155.1	5.7	266.7	44

Table 1.2 – The quantity (tonnes) of lettuce imported to and exported from the UK between 2005 and 2009. (Defra Horticultural Statistics, 2009)

1.3.5 Lettuce physiology, growth and development

Lettuce leaves are spirally arranged on a short stem and form a rosette of leaves. The rosette growth may continue throughout the vegetative period of growth of the plant, or may form a round head, as with butterhead/crisphead varieties, or an elongated head may develop as with cos type lettuces. The earliest leaves to develop are elongated and gradually become broader with subsequent leaf formation.

Lettuce is an annual crop, with the reproductive phase of developmental growth starting when vegetative growth reaches a mature stage. Usually a single stem emerges from the lettuce plant which then forms axillary buds. The inflorescence produced is a corymbose panicle, which is composed of many capitula or flower heads, each capitulum consists of several florets (Ryder, 1999). The florets are surrounded by three to four rows of bracts, which form an involucre. Each floret

consists of a single, yellow petal with five teeth; the lower part is fused as a tube and surrounds the sexual parts. Each of the florets has a double carpel, consisting of an elongated style and a divided stigma, there are five stamens and the anthers are fused forming a tube. Flowers open just once a day, in the morning, as they open the style elongates while the anthers dehisce from within and the shed pollen is swept upwards by the style and stigma hairs. A fertilised ovary forms an embryo which is surrounded by nuclear and endosperm tissue and a thin pericarp; known as an achene. Hair like pappus develops on top of the achene; this matures two weeks after fertilisation.

Due to the structure of the flower, lettuce is an obligate self-fertilising species, the elongation of the style occurs at the same time as the pollen is released from the inner surface of the fused anther tube. The stigmas and style get covered with pollen when they emerge; the grains then germinate and penetrate the stigmata surface very quickly, figure 1.1 shows the stages of anthesis of a lettuce floret. Cross-pollination between lettuce plants occurs at a very low rate; approximately only 1 % (Thompson *et al*., 1958), this is because the pollen grains are very sticky, and do not get carried by the wind and also because few insects tend to work lettuce flowers.

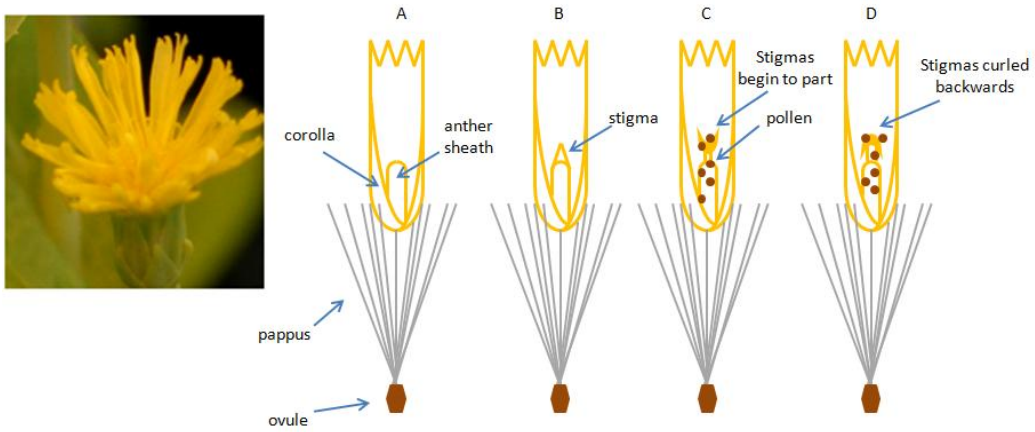


Figure 1.1 – Stages of anthesis of a lettuce floret – (A) Stigma has not emerged from anther sheath (B) Stigma emerges (C) Stigmas, covered with pollen begin to part; ideal stage for crossing (D) Stigmas curl backwards; pollination is complete; flowers are past the ideal crossing stage (Adapted from Ryder 1986; Ryder, 1999)

1.3.6 The lettuce genome

L. sativa is a diploid (2n=18) species, (Koopman and De Jong 1996), with an estimated genome size of 2.6 Gb (Michaelson *et al*., 1991), which is approximately 16 times the size of *Arabidopsis* (0.16 Gb). The wild *Lactuca* species; *L.serriola*, *L. saligna* and *L. virosa* also have 18 chromosomes (Doležalová *et al*, 2002). All four are obligate self-fertilising species.

The genetic differences between *L. sativa* and the three most common wild *Lactuca* species is still unclear. Analysis of the variation in chromosome number and the nuclear DNA content of the *Lactuca* species by Doležalová *et al*, (2002), showed that when compared to data in vascular plants with a known genome size (Bennett and Leitch, 1997), *Lactuca* belongs to a group of species with a medium genome size. Different relative DNA amounts were found in *L. sativa*, *L. serriola*, *L. saligna* and *L. virosa*. The largest relative DNA content was observed in *L. virosa* accessions and the lowest was seen in *L. saligna* (Doležalová *et al*, 2002). This agrees with previous work completed by Koopman and De Jong (1996). Dendrograms constructed based on this information, along with statistical analysis showed that *L. saligna* and *L. virosa* lie in groups far removed from a group containing *L. sativa* and *L. serriola* (Doležalová *et al*, 2002). This data is further backed up by the RFLP work carried out by Kesseli *et al*., (1991), but is in conflict with the results of the chromosome banding patterns seen Koopman *et al*., 1993, which suggest the close evolutionary distance between *L. sativa*, *L. serriola* and *L. saligna*, with *L. virosa* being more distant and the work carried out by Zohary, (1991) which suggests that *L. saligna* belongs to a secondary *Lactuca* gene pool and *L. virosa* to a tertiary gene pool. These studies provide more evidence for *L. serriola* being the progenitor of *L.*

sativa, and confirm that *L. virosa* is the most evolutionary different to cultivated lettuce.

The lettuce life cycle can vary in length from a few weeks in mid-summer for butterhead varieties to a few months for crisphead/cos varieties. Despite the cultivars of *L. sativa* being highly inbred crosses can be made readily and multiple generations can be produced each year.

A typical breeding program to create a new lettuce variety takes around 10-12 years. Salinas, a crisphead variety took 13 years (1962-1975) to breed (Ryder 1986), the process involves not only breeding stages of development, but also selection processes and a number of trials with breeders and growers which eventually lead to the release of the new cultivar.

1.3.7 Flowering in lettuce

The ability to use genetics to improve the control of flowering time of a crop offers potentially great advantages as it would enable more consistent crop performance and reduce seasonal variation. Bolting corresponds to the transition from vegetative to the reproductive phase (i.e. flowering) in vegetable crops. The time of bolting in lettuce is agronomically important because it impacts on quality, yield and scheduling of production (Ryder 1986 ; Nishioka *et al*., 2005).

Most plants have a juvenile phase which prevents floral induction until a certain developmental stage has been reached and the plant is of a sufficient size to ensure resources are available to sustain flower production. A juvenile phase length has not been identified for lettuce, but is thought to exist due to the lack of response to environmental cues (photoperiod and temperature) in young lettuce plants. The length of any juvenile phase is also likely to vary depending on the lettuce variety.

Lettuce, like *Arabidopsis* is a quantitative long day (LD) plant, flowering earlier under photoperiods of 16h, with some varieties also responding quantitatively to vernalisation (Prince *et al*., 1979). Bolting is also promoted by higher temperatures; an increasing problem for growers with the current trend for warmer summers (Poh *et al*., 1998; Ryder and Waycott, 1994; Ryder, 1999). Lettuce is also believed to have a light quality requirement, some varieties will not mature under some artificial lighting, e.g. high pressure sodium bulbs, a blue light requirement is thought to be key to this response (Professor Thomas; Professor Pink, University of Warwick, personal communication). Bolting is also promoted by nitrogen fertiliser (Soundy and Smith, 1994) and gibberellin levels (Waycott and Ryder, 1993). Lettuce plants that are in the early stages of bolting are visibly indistinguishable from non-bolting plants, however there are changes in the production of secondary metabolites such as sesquiterpene lactones and phenolic compounds, including lactucin, deoxylactucin and lactucopicrin (Sessa *et al*., 2000) which are produced to protect the young floral bud from insect attack. These compounds give the plant a bitter taste which renders the crop unsaleable (Price *et al*., 1990; Bennett *et al*., 2002). Resistance to bolting or 'holding ability' in the field in lettuce is an important quality character, (Pink and Keane 1993; Ryder and Milligan 2005), as it preserves the quality of lettuces sold for consumption, and increases the sustainability by reducing crop losses and wastage which industry representatives have identified as an area requiring further research (http://www.plbr.cornell.edu/psi/Organic%20Breeding%20

Roundtable%20Summaries%202005.pdf).

Bolting is also problematic in other crop species including chicory, sugar beet, rocket, radish, turnip and spinach. Relatively little research into delaying bolting in these species has been carried out. An EMS mutagenised sugar beet population has

produced five non-bolting lines (Hohmann *et al*., 2005). It was shown that an antisense copy of the *Arabidopsis GIGANTEA* (*GI*) gene caused delayed bolting and flowering in radish (Curtis *et al*., 2002). QTLs have been mapped for bolting time in turnip under different environmental conditions; interestingly bolting is promoted by a decrease in temperature as well as an increase in photoperiod (Nishioka *et al*., 2005). The bolting time in chicory was investigated at low temperatures by Dielen *et al*., (2005), the bolting phenotypes varied depending on the variety, age of plant and duration of the vernalisation period.

1.4 Flowering pathways in the model plant *Arabidopsis*

1.4.1 Introduction

In many plants, the transition from vegetative growth to flowering is controlled by both endogenous and environmental cues, which are difficult to predict and/or control (Corbesier and Coupland 2005). One of the key factors contributing to the success of flowering plants is their ability to regulate the timing of flowering so they can take maximum advantage of the most environmentally favourable conditions to ensure survival of their progeny (Devlin and Kay 2000).

Flowering has been characterised at the genetic level in the model plant, *Arabidopsis thaliana*. Studies have identified it as a facultative LD plant and as such will flower in both long and short days (SD), but much earlier in LD conditions. In addition to this photoperiod response pathway, other pathways also regulate flowering time in *Arabidopsis*; these are the vernalisation, autonomous, gibberellin (GA), light quality, and ambient temperature pathways. Variation in environmental conditions will dictate which of the different *Arabidopsis* flowering pathways has the major influence on flowering time (Simpson and Dean 2002).

The predominance of the different pathways also changes with the developmental state of the plant. Early on in the life cycle of the plant flowering is actively repressed to enable the plant to grow sufficiently large to be able to support the development of flowers, fruits and seeds (Thomas *et al*., 2006). This repression is gradually reduced as the plant develops by the floral enabling pathways, which include the vernalisation and autonomous pathways (Boss *et al*., 2004). Under certain environmental conditions floral promotion pathways are activated, this includes the photoperiodic, GA, ambient temperature and light quality pathways (Thomas *et al*., 2006), see figure 1.2. Genes that play a role in these flowering pathways have been identified in *Arabidopsis* through the analysis of natural variation in different ecotypes and through the characterisation of induced mutations (Levy and Dean 1998).

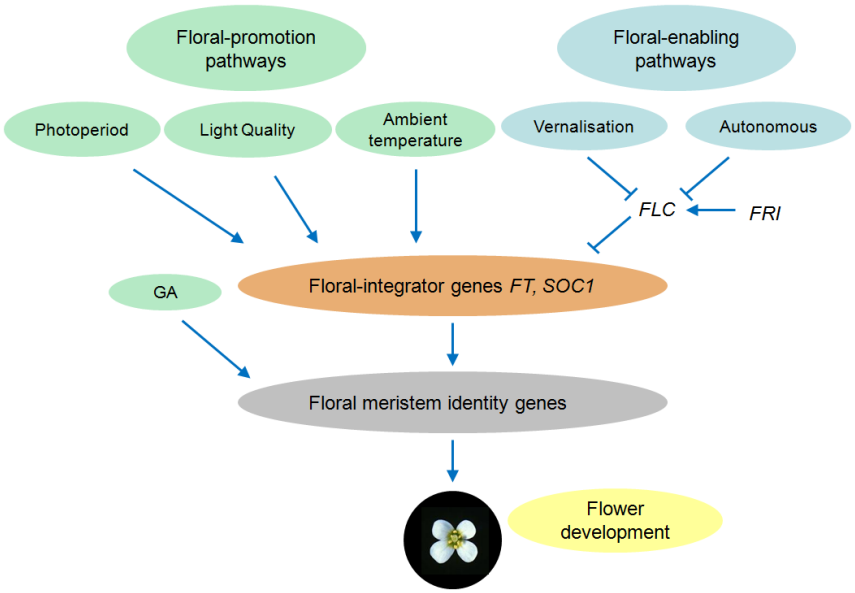


Figure 1.2 – Integration of the environmental and developmental pathways leading to flowering in *Arabidopsis*. (Adapted from Thomas *et al*., 2006)

1.4.2 The photoperiodic pathway

Arabidopsis shows a strong photoperiodic response in the onset of flowering, and most ecotypes flower in spring or early summer as the days become longer. In

laboratory conditions, flowering occurs much earlier under LD of 16h light than under SD of 10h light (Searle and Coupland 2004). Among the flowering-time mutants of *Arabidopsis*, there is a group that are late flowering in LD but which flower at a similar or identical time to wild type (WT), under SD. These mutants are weakly, or not at all, sensitive to vernalisation, and were proposed to define a genetic pathway that promotes flowering specifically in response to LD called the photoperiodic pathway (Koornneef *et al*., 1991).

The photoperiodic pathway detects and responds to the duration of the daily light period. Light signals are detected by photosensory receptors including the red/far-red light-receptor phytochromes (*PHYA-E*) and the blue/UV-A light-receptor cryptochromes (*CRY1* and *2*) (Quail 2002). These receptors mediate light-regulated plant growth and development from seed germination to flower initiation (Guo *et al*., 1998). The photoreceptors interact to entrain the circadian clock to a 24 h period. The circadian clock is an endogenous timekeeping mechanism which is controlled by several negative feedback loops (Halliday *et al*., 2003; Jackson, 2009), enabling the clock to cycle in constant conditions, without entrainment by zeitgeber (ZT) signals, such as changes in light or temperature. ZT refers to any exogenous cue that synchronises an organism's endogenous time-keeping system to a 24 h light/dark cycle. The circadian clock co-ordinates a number of processes including photosynthesis and floral initiation (McClung, 2006).

The negative feedback loop involves the *TIMING OF CAB EXPRESSION 1* (*TOC1*) gene which positively regulates two partially redundant Myb transcription factors; LATE ELONGATED HYPOCOTYL (LHY) and CIRCADIAN CLOCK ASSOCIATED 1 (CCA1). These proteins feed back to negatively regulate the expression of *TOC1* by binding to an evening element in its promoter (Alabadi *et al*.,

2001). *ZEITLUPE* (*ZTL)* also regulates *TOC1* by targeting the protein for degradation via the 26S proteosome (Mas *et al*., 2003), see figure 1.3. A further gene *FIONA1* (*FIO1*), has recently been identified as having a role in regulating the period length of the clock in close association with the central oscillator (Kim *et al*., 2008).

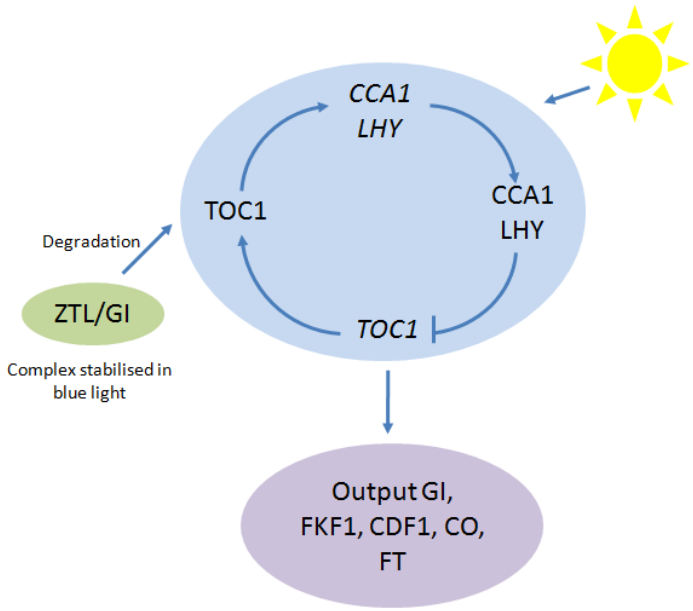


Figure 1.3 – Simplified version of the circadian clock. The expression of the *TOC1* gene is regulated by a negative feedback loop which positively regulates two partially redundant Myb transcription factors; LHY and CCA1. These proteins feed back to negatively regulate the expression of *TOC1* by binding to an evening element in its promoter. *ZTL* further regulates *TOC1* by targeting the protein for degradation. *TOC1* promotes the expression of *GI* which in turn forms a complex with *FKF1* in blue light which acts to regulate *CO* expression through *CDF1*. *CO* shows a strong circadian peak at the end of a LD; *CO* is stabilised by *CRY2/PHYA*. *CO* promotes the floral integrator *FT* which promotes the floral meristem identity genes resulting in flowering.

PHYB is the major photoreceptor for normal levels of red light and *PHYA* is important for perception of far-red light and in the Very Low Fluence Response (VLFR) and High Irradiance Response (HIR) to red light. *CRY1* and *CRY2* are blue light photoreceptors affecting flowering, they act to stabilise the CONSTANS (CO) protein at the end of a LD (Valverde *et al*., 2004). *CO* is expressed at the end of a LD and CO protein is degraded at night. *CO* regulates the floral integrator gene

FLOWERING LOCUS T (*FT*) which in turn promotes expression of *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) which leads to floral initiation (Samach *et al*., 2000; Yoo *et al*., 2005), see figure 1.4.

Flowering will occur when *CO* expression and a blue or far-red light signal occur simultaneously. Light is important in controlling the stability of the CO which is promoted by blue and far-red light. In the dark, *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*) promotes the degradation of CO (Jang *et al*., 2008) leading to a delay in flowering in SDs. CO is also degraded in the morning, this occurs through *PHYB* (Valverde *et al*., 2004).

Light is also important in promoting the formation of a complex between GIGANTEA (GI) and FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1). This complex is specifically formed under blue light conditions (Sawa *et al*., 2007). *CO* expression is regulated by *GI* and *FKF1*, they regulate *CO* by combining to form a complex which directly regulates *CYCLING DOF FACTOR 1* (*CDF1*) stability, resulting in its repression (Sawa *et al*., 2007). *CDF1* is a repressor of *CO.*

More details on the role of *CO* and *FKF1* including details on gene structure is described in sections 3.3.10 and 3.3.12.

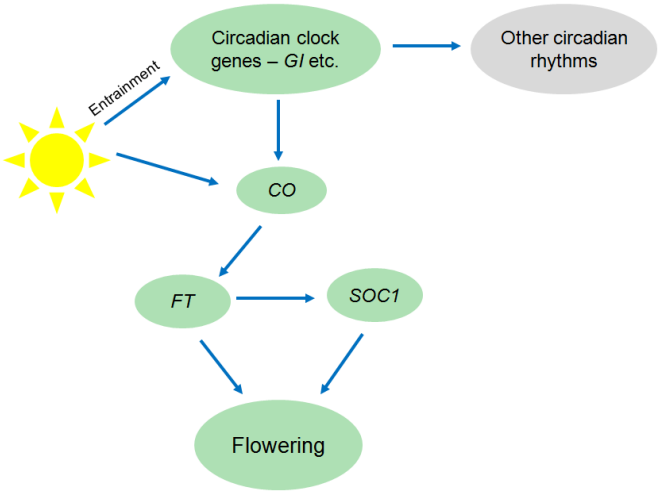


Figure 1.4 - Model of the LD photoperiodic pathway. (Adapted from Suárez-López *et al*., 2001). In *Arabidopsis* when high levels of CO protein coincide with the light period, *FT* expression is induced resulting in flowering. This occurs in LD, but not in SD when CO protein levels do not rise high enough during the light period to induce *FT* expression (Samach *et al*., 2000; Suárez-López *et al*., 2001).

MicroRNAs (miRNA) also play a role in regulating photoperiodic flowering in *Arabidopsis*. Jung *et al*., (2007) showed that miR172 regulates photoperiodic flowering through *FT*, independently of *CO*. miR172 is regulated by *GI*, this suggests a new sub-pathway involved in controlling flowering via photoperiod. It would appear that there are still more genes to be discovered and characterised that are involved in controlling flowering through the photoperiodic pathway, suggesting a more complex genetic pathway exists.

1.4.3 Vernalisation pathway

Both the vernalisation and autonomous pathways converge on the floral repressor *FLOWERING LOCUS C (FLC)*, see figure 1.2. Vernalisation is the acceleration of flowering by exposure to prolonged cold. A vernalisation period can range from as little as four weeks to a number of months at temperatures between 4-8 °C. Such a requirement for vernalisation is associated with a winter annual growth habit (Napp-Zinn, 1987). Many *Arabidopsis* ecotypes will flower very late if not exposed to a vernalisation treatment; however they will flower early if subjected to a vernalisation

treatment (Michaels and Amasino, 2000). *FLC* and *FRIGIDA (FRI)* genes are required to confer a vernalisation requirement (Johanson *et al*., 2000; Clarke and Dean 1994). *FRI* promotes high levels of *FLC, FLC* in turn represses *FT*, and *SOC1*, resulting in flowering being delayed (Michaels and Amasino 1999). *FLC* mRNA levels fall when plants are exposed to cold.

Considerable variation in the vernalisation requirement and flowering time between different *Arabidopsis* accessions has been observed. This variation is linked to the geographical region associated with the individual accessions; this supports the idea that variation is important for adaptation (Shindo *et al*., 2005). Once *FLC* expression is reduced by vernalisation it is stably maintained by an epigenetic mechanism as plants continue growing through the spring and summer. Mutant screens of accessions unable to suppress or epigentically control *FLC* expression have identified proteins which mediate vernalisation; *VERNALISATION 2 (VRN2)* (Gendall *et al*., 2001), *VERNALISATION 1 (VRN1)* (Levy *et al*., 2002) and *VERNALISATION INSENSITIVE 3 (VIN3),* a PLANT HOMEODOMAIN (PHD)-finger containing protein (Sung and Amasino, 2004). VRN2 is homologous to the *Drosophila melanogaster* polycomb group protein Suppressor of Zeste 12 (Wood *et al*., 2006). In plants VRN2 is part of a Polycomb Repressive Complex like complex which acts to regulate *FLC* expression in both non-vernalised and vernalised conditions (Wood *et al*., 2006).

VIN3 modifies the normally open chromatin structure of *FLC* by deactylation of histones; this closes the chromatin and *VRN1* and *VRN2* stably repress the expression of *FLC*. It has been shown that in accessions which require a long period of vernalisation that *FLC* expression was reactivated following nonsaturating vernalisation; this reactivation showed a correlation to the rate of accumulation of

FLC histone H3 Lys 27 trimethylation (H3K27me3) (Shindo *et al*., 2006). This suggests that the epigenetic control of *FLC* explains the adaptation of *Arabidopsis* accessions to different geographical areas.

Vernalisation can overcome the repressive effects of *FLC* and *FRI* and can also compensate for the lack of any autonomous pathway genes. The vernalised state is reset at meiosis to ensure that the vernalisation requirement for flowering is present in the next generation. *FLC* reprogramming is composed of three phases; it is repressed in gametogenesis before being reactivated in early embryogenesis and maintained in late embryogenesis (Choi *et al*., 2009). The *FLC* gene structure and function is discussed in more detail in section 3.3.13.

1.4.4 Autonomous pathway

The seven genes making up the autonomous pathway; *FCA*, *FLD* (*FLOWERING LOCUS D*), *FLK* (*FLOWERING LOCUS KH DOMAIN*), *FPA*, *FVE*, *FY* and *LD* (*LUMINIDEPENDENS*) act in parallel to the vernalisation pathway to promote flowering by repressing the expression of the floral repressor *FLC* (Koornneef *et al*., 1991; Marquardt *et al*., 2006). Mutations in genes in the autonomous pathway result in elevated *FLC* levels and delayed flowering (Michaels and Amasino 2001).

The autonomous pathway was identified through the isolation and characterisation of a collection of late flowering *Arabidopsis* mutants (Koornneef *et al*., 1991). Four of the mutants *fca*, *fy*, *fpa* and *fve* flowered later than WT in both LD and SD photoperiods (Koornneef *et al*., 1998), responses that are similar to those seen in other late-flowering mutants; *ld* (Lee *et al*., 1994), *fld* (Sandra and Amasino, 1996) and *flk* (Lim *et al*., 2004). The late flowering mutants retain a photoperiod response and therefore are thought to promote flowering in a photoperiod-independent

pathway. Unlike photoperiod pathway mutants, the late-flowering phenotype of autonomous pathway mutants is strongly suppressed by vernalisation. Thus, the late flowering, vernalisation-responsive phenotype of autonomous pathway mutants is similar to the winter annual habit of *FRI*-containing late-flowering ecotypes.

FCA (MacKnight *et al*., 1997), FPA (Schomburg *et al*., 2001) and FLK (Lim *et al*., 2004) have all been shown to interact with RNA. All three encode plant specific proteins which are predominantly found in the nucleus. FCA and FPA both encode RNA-recognition motif (RRM)-type RNA-binding proteins. FCA has two RRM domains along with a WW protein interaction domain (MacKnight *et al*., 1997), FPA contains three RRMs (Schomburg *et al.*, 2001) and FLK contains three K-homology (KH)-type RNA binding domains (Lin *et al*., 2003). FY interacts with the WW domain of FCA and is required for FCA to promote flowering (Simpson *et al*., 2003). FY is highly conserved to *S. cerevisiae* protein Pfs2p which cleaves and polyadenylates mRNA. LD encodes a homeodomain protein (Lee *et al*., 1994) a domain which is most often associated with DNA binding, in some cases it has been reported to bind to RNA (Simpson, 2004). *FLD* and *FVE* regulate *FLC* epigenetically. FLD is a homologue of the human protein KIA0601 (He *et al*., 2003), both of which are similar to polyamine oxidases and contain the SWIRM (SWI3p, Rsc8p, Moira) domain. This domain is found in a range of enzymes that are involved in chromatin remodeling. FVE is a plant homologue of a yeast protein; MSI (MULTICOPY SUPPRESSOR OF IRA1). Both genes contain WD-repeat proteins that are involved in chromatin assembly or modification. FLD and FVE are required for histone deacetylation of *FLC* chromatin (He *et al*., 2003; Ausin *et al*., 2004). Figure 1.5, a schematic adapted from Simpson (2004) shows the basic actions

of the autonomous pathway genes. Each of the genes is discussed in more detail in sections 3.3.3–3.3.9.

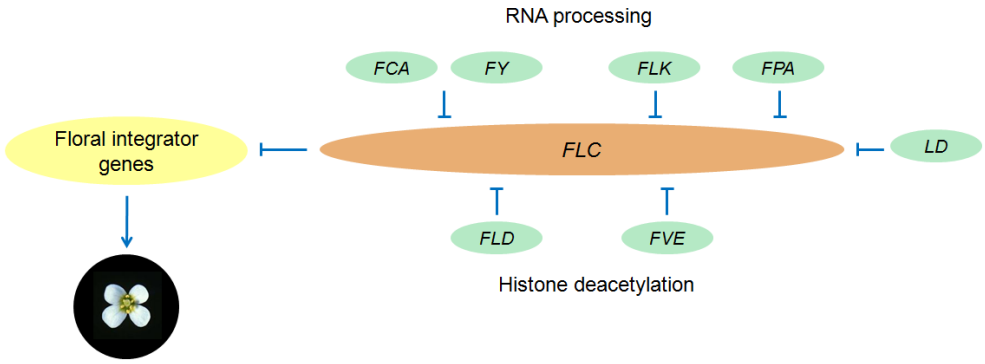


Figure 1.5 – The autonomous pathway and its role in flowering. Figure adapted from Simpson, 2004.

The autonomous pathway is relatively understudied, but recent research has shown further roles and potential interactions between *FCA*, *FPA* and *FLD*. Liu *et al*., (2007) have shown that *FCA* requires *FLD* for *FLC* downregulation. This interaction occurs in one of two ways; either FCA function is involved in RNA processing with the antisense transcripts leading to FLD induced histone modifications, or FLD/FCA induced chromatin modifications result in altered FLC transcript processing. The function of FCA also partially depends on the *DICER-LIKE 3 (DCL3)* gene, which is involved in chromatin silencing; a number of small interfering RNAs (siRNAs) are also involved in localising and maintaining these chromatin modifications (Bäurle *et al*., 2007). Liu et al., (2010) have further shown that the targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. Bäurle *et al*., (2007) also proposed that *FCA* and *FPA* regulate chromatin silencing of single and low-copy genes and interact in a locus dependant manner with the siRNA-directed DNA methylation pathway to regulate common targets. Further work has speculated that FCA and FPA protein functions as part of a transcriptome surveillance mechanism linking RNA recognition with chromatin silencing mechanisms (Bäurle and Dean, 2008). They suggest that *FCA* and *FPA* act to promote *FLD* which in turn

represses *FLC* expression; figure 1.6 proposes a more updated version of how part of the autonomous pathway may be interacting before converging to repress *FLC*.

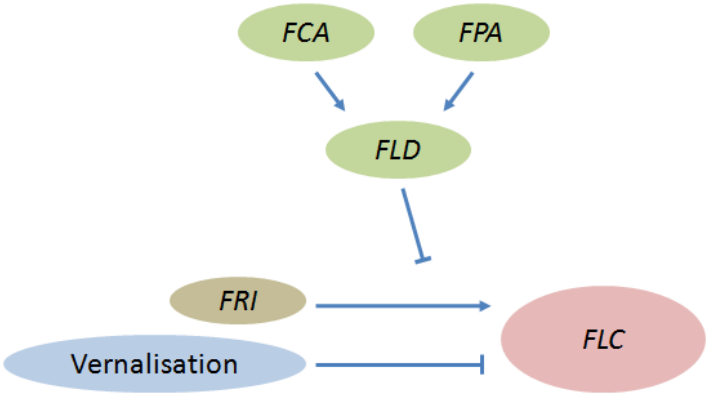


Figure 1.6 – A new role for *FCA*, *FPA* and *FLD* within the autonomous pathway. *FCA* and *FPA* act through *FLD* to repress *FLC*. Adapted from Bäurle and Dean, 2008.

In further research by Veley and Michaels (2008) it has also been shown that the autonomous pathway genes play roles in non-flowering related functions which are independent of *FLC*, for example *FPA* plays functionally redundant roles with *FLD*, *FVE* and *LD* in plant growth and development.

It was widely believed that the expression of the autonomous pathway genes are not regulated by any other flowering time gene, it has however been shown that *AGAMOUS-LIKE 28* (*AGL28*), a gene involved in cellular differentiation during fruit and leaf development (Gu *et al*., 1998) promotes flowering by promoting the expression of the autonomous pathway genes (Yoo *et al*., 2006). The autonomous pathway is a complex one, the understanding of which is slowly beginning to take shape.

Autonomous pathway mutants in lettuce may be of greatest interest commercially as the late flowering effect of these mutations can be overcome by a vernalisation treatment, enabling flowering to take place normally when seed production is required which is of importance to plant breeders, see figure 1.7.

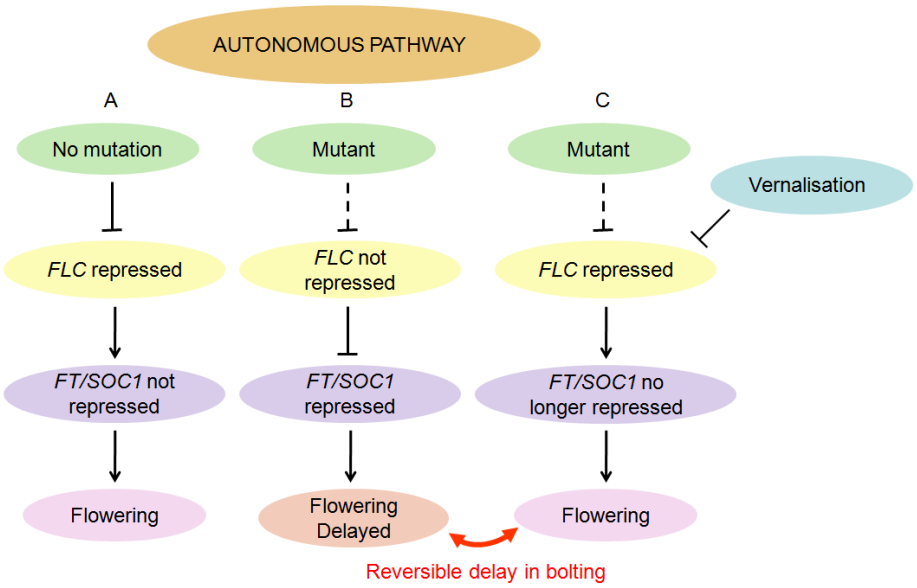


Figure 1.7 – Potential manipulation of the Autonomous pathway to benefit both lettuce growers and breeders. Autonomous pathway genes act through *FLC* to affect flowering (A). Autonomous pathway mutants delay flowering as they no longer act to repress *FLC*, *FLC* in turn represses the floral integrators which results in later flowering (B). However, delayed flowering, caused by mutations in the autonomous pathway can be overcome by a vernalisation treatment (C).

1.4.5 Gibberellin, light quality and ambient temperature pathways

The growth hormone, GA promotes flowering in *Arabidopsis*. This was initially demonstrated by applications of exogenous GA (Langridge 1957) and more recently by using mutations that disrupt either GA biosynthesis or signaling which delay flowering, especially in SDs (Wilson *et al*., 1992). Plants that over-express *GA-20 oxidase*, a gene involved late on in the GA pathway, are early flowering in both LDs and SDs (Huang *et al*., 1998; Coles *et al*., 1999).

More recently two other pathways affecting flowering time in *Arabidopsis* have been identified. One is the light quality pathway where light quality affects *FT* expression in a *CO*-independent manner (Cerdan and Chory 2003). The other is the ambient temperature pathway where *FT* and *SOC1* expression is generally much higher at 23°C than 16°C (Blázquez *et al*., 2003). Blázquez also showed that autonomous pathway mutants flower at the same time regardless of ambient temperature.

1.4.6 Floral integrators

The photoperiodic, vernalisation, autonomous, ambient temperature and light quality pathways converge on the floral integrator genes *FT* and *SOC1* which activate the floral meristem identity genes, *LEAFY (LFY)* and *APETALA1 (AP1)*, whose actions convert the shoot apical meristem to a reproductive fate (Boss *et al*., 2004). Corbesier *et al*., (2007) and Tamaki *et al*., (2007) concluded that FT protein acts as a long-distance signal that moves from leaf to apex to induce flowering in *Arabidopsis* and rice. *FT* is described in more detail in section 3.3.2.

1.5 Conservation of flowering genes in crop species

Key components within the pathways that affect flowering in *Arabidopsis* have been identified, however to date there has been little, if any, translation of this knowledge to achieve real benefits in crop plants.

Research has shown that sequence synteny exists between *Arabidopsis* and other crop species, including monocotyledonous plants. Comparative analyses with the rice genome sequence, (a SD plant), revealed that a majority of the *Arabidopsis* key regulatory genes for flowering time are conserved. Orthologues of *GI* (*OsGI*), *CO* (*HEADING DATE 1*; *Hda1*) and *FT* (*Hd3a*) have been identified in rice. *OsGI* promotes *Hd1* expression, as occurs in *Arabidopsis* with *GI* and *CO*. However, *Hd1* inhibits *Hd3a*, the rice *FT* orthologue in LDs, but promotes its expression in SDs; therefore promoting flowering in SDs (Kojima *et al*., 2002; Izawa *et al*., 2003; Hayama *et al*., 2003). Lee *et al*., (2005) also showed that a rice homologue of the autonomous pathway gene *FCA* is able to partially rescue the late flowering phenotype of the *Arabidopsis fca* mutant.

Five genes highly conserved to *FT* have been identified in barley, a LD plant which is phylogenetically closer to rice than *Arabidopsis*. *HvFT1* was identified as the main barley *FT*-like gene involved in the switch to flowering (Faure *et al*., 2007). A barley *GI* (*HvGI*) and two *CO* (*HvCO1* and *HvCO2*) genes have also been reported (Griffiths *et al.*, 2003; Dunford *et al*., 2005). The major component of flowering in LD photoperiods in barley is the pseudo-response regulator, *PHOTOPERIOD-H1* (*Ppd-H1)* (Decousset *et al*., 2000). A reduced photoperiod response is seen in the *ppd-h1* mutant (Turner *et al*., 2005). This can be explained by an altered circadian expression of *CO* and a reduced expression of *FT*. Reduced photoperiod response is highly desirable in barley, so this gene is of particular interest. Interestingly this highlights that some differences exist between species in the genetic control of the photoperiod response.

Work to find pea orthologues of *Arabidopsis* flowering time genes identified a number of genes which are conserved with *CO*, *FT* and *FCA*, as well as a *GI* orthologue; *LATE BLOOMER 1* (*LATE1*) (Hecht *et al*., 2005; Hecht *et al*., 2007). Four *FT* paralogs have been identified in sunflower (Blackman *et al.*, 2010) and an *FT* orthologue along with two *CO* homologues have been identified in grape (Carmona *et al.*, 2007; Almada *et al*., 2009). The large scale comparison of the flowering time pathways between *Brachypodium distachyon*, a model species for the temperate grasses, with rice and *Arabidopsis* has shown large scale conservation of the components which initiate the floral transition. The results suggest that an ancient core photoperiodic pathway that promotes flowering via the induction of *FT* has been modified by the recruitment of additional lineage specific pathways that promote or repress *FT* expression (Higgins *et al*, 2010).

Potentially of interest to this project is the work done regarding the floral transition in sugar beet, bolting in this crop, like with lettuce is undesirable. Bolting in sugar beet is strongly linked to the early bolting (*B*) locus. A *CO* orthologue has been identified through the functional complementation of the *Arabidopsis co-2* mutant. Interestingly, this gene does not map to the *B* locus, in addition, it shows a different expression pattern to *Arabidopsis CO* suggesting a different photoperiodic response in sugar beet (Chia *et al*., 2008).

Significant levels of local synteny have been detected at a fine scale covering segments of 1Mb regions of *Arabidopsis* and regions of less than 5cM in lettuce and sunflower (Timms *et al*., 2006).

1.6 Lettuce Resources

1.6.1 EST databases

A number of useful resources exist for lettuce that are of use for this project. Lettuce expressed sequence tag, (EST), homologues of *Arabidopsis* sequences have been isolated by Professor Michelmore's group at the University of California at Davis (UC Davis), California. Much of this data is accessible through the Compositae Genome Project (CGP), database, (http://compgenomics.ucdavis.edu/ index.htm). The EST libraries were made from ten pools of RNA comprising of different tissues/developmental stages/environmental conditions from the two lettuce species *L. sativa* cv. Salinas and *L. serriola*. The first phase of the CGP, (CGP1), generated over 68,000 ESTs (Michelmore 2007). Data released in early 2007 represented the second phase of CGP, (CGP2), this included a further 160,000 lettuce ESTs. These ESTs were generated from *L. sativa* cv. Salinas, *L. serriola, L. saligna L. virosa* and *L. perennis.* Overall, the ESTs have been compiled into 35,000

unigenes, which are displayed on the CGP website (http://cgpdb.ucdavis.edu/), representing an estimated 80 % of the genes in lettuce. A collaboration between our research group and Professor Michelmore was agreed in order to share data and resources.

1.6.2 Genomic resources

A *L. sativa* cv. Diana BAC library representing two to three genome equivalents has been constructed. The library comprises over 50,000 clones with an average insert size of 110 kb (Fritjers *et al*., 1997). A mapping population of 300 RILs (Agryris *et al*., 2005; Truco *et al*., 2007) and an integrated map comprising over 2,700 markers in nine linkage groups exists for lettuce, (Kesseli *et al*., 1994), and QTLs have been identified for a large number of traits including flowering time; these are currently being mapped to candidate genes by Professor Michelmore's group.

An Affymetrix high density GeneChip® microarray has been developed for massively parallel genetic mapping and sensitive expression analysis. The chip has a total of 6.6 million features representing ~35,000 unigenes based on EST sequences, each unigene is represented by a 500 bp oligonucleotide which has a 2 bp stagger (http://chiplett.ucdavis.edu/index.php). Over the next few years efforts are being made to use the new generation of sequencing technologies to obtain whole genome sequence for lettuce.

1.6.3 Plant resources

1.6.3.1 EMS mutagenised populations

Two EMS mutagenised populations have been produced at Warwick HRI by Professor Pink's group. EMS randomly induces point mutations throughout the

genome of the treated seed and primarily causes G:C to A:T transitions (Krieg, 1963). Research carried out using an EMS treatment on lettuce seeds has previously produced chlorophyll and leaf shape mutants (Robinson, 1986) suggesting the species is receptive to the treatment. The first TILLING population is in the butterhead lettuce cv. Larissa background; the other is in the crisphead lettuce cv. Saladin see figure 1.8. The Larissa population formed a pilot study for the work and as such is relatively small. The creation of the populations is described in section 4.3.1 and 4.4.1. Larissa has a relatively rapid life cycle and was primarily bred for protected cropping over the SD winter months. Saladin was chosen to create a large scale EMS population as it is widely regarded as the progenitor of most modern day lettuce varieties.

Figure 1.8 – Lettuce cultivars used for the EMS TILLING populations. *L. sativa* cv. Larissa, (left) and *L. sativa* cv. Saladin, (right).

Seed treated with EMS can be screened for a particular trait of interest, in this case the number of days to bolt. Genes involved in flowering time within plants displaying a LB phenotype compared with WT plants can be screened at the molecular level to identify sequence differences.

One such screening method is TILLING, (Targeting Induced Local Lesions IN Genomes), it is an effective reverse genetic technique used to screen populations mutagenised with chemical mutagens. TILLING involves heteroduplex analysis to

detect which organisms in a population carry mutations within specific genes (McCullum *et al*., 2000; Till *et al*., 2003), see figure 1.9. TILLING populations have been successfully created in a number of crop species including Maize (Till *et al*., 2004), Rice (Till *et al*., 2007), Wheat (Slade *et al*., 2005), Pea (Dalmais *et al*., 2008) and Tomato (Minoia *et al*., 2010). This method was used successfully on a population of EMS mutagenised sugar beet to identify non-bolting mutants (Hohmann *et al*., 2005). An adaptation of this method, Ecotilling enables the discovery of polymorphisms in natural populations (Comai *et al*., 2004).

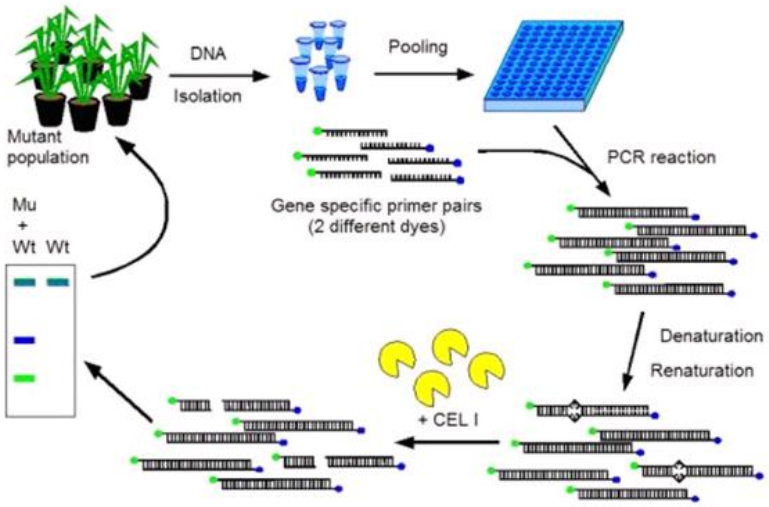


Figure 1.9 – High throughput TILLING of *Arabidopsis.* Seeds are mutagenised with EMS. To avoid ambiguities caused by chimerism of mutant plants in the first (M1 generation), they are self-fertilised, and M2 progeny from single seed descent are used for screening. Tissue is collected from the M2 generation and DNA is extracted. DNA is pooled and PCR is performed using 5' end labeled gene specific primers to target loci. Heteroduplexes are formed by heating and cooling PCR products. CEL I nuclease is used to cleave at base mismatches, and different length products caused by induced mutations are visualized with denaturing polyacrylamide gel electrophoresis. Figure taken from www.risoe.dk/rispubl/PRD/AR2003/plr_Eco-Tilling.htm.

1.6.3.2 Retrotransposon mutagenised lines

Mazier *et al*., (2007) described the successful use of the 5.3 kb-long *Tnt1* retrotransposon from tobacco for gene tagging in lettuce. They showed that the retrotransposon could be mobilised by an *in vitro* tissue culture regeneration step but that it was otherwise stable in subsequent generations. Seeds from T1 generations obtained from selfing T0 primary transformants, and T2 generations obtained on

selfing T1 plants, were made available from the lettuce cultivar Jessy for screening for bolting time. Nine lines were made available for analysis which included WT, one T1 and seven T2 Jessy lines.

1.6.3.3 Wild lettuce diversity set

Table 1.3 details the 120 lines from three wild lettuce species that have been selected to make up a wild lettuce diversity set that is representative of a range of bolting times. The set includes previously identified LB *L. virosa* lines from CGN, Netherlands. For the accession names of the lines screened see Appendix I, table A1.

Seed Source	Species	Number
Genetic Resources Unit, Warwick Horticultural International, UK	*L. saligna*	8
	L. serriola	27
	L. virosa	15
Centre for Genetic Resources, Wageningen, NL	*L. virosa*	4
University of California at Davis, USA	*L. saligna*	34
	L. serriola	20
	L. virosa	12

Table 1.3 – Sources of the wild lettuce diversity population studied.

1.7 Project strategy

The project was split into three main research areas, each made up of very specific aims. These research areas make up the three results chapters reported in this thesis.

1.7.1 Isolation of flowering time gene homologues from lettuce

The first area of research focused on obtaining candidate genes that were likely to be involved in the regulation of flowering time in lettuce. To identify these genes from lettuce an applied approach was undertaken based on the previously characterised genes involved in well defined flowering pathways in the model plant species *Arabidopsis*. Candidate genes involved in regulating flowering time in *Arabidopsis*

were targeted in lettuce; the *Arabidopsis* gene sequences used in this project were obtained from The *Arabidopsis* Information Resource (TAIR). The 12 key flowering time genes targeted are shown in table 1.4; they were chosen using a number of criteria.

Gene	*Arabidopsis* Locus	Flowering Function
FLC	AT5G10140	Floral repressor Convergence point for autonomous and vernalisation pathways
CO	AT5G15840	Important role in photoperiodic pathway
CRY2	AT1G04400	Blue light photoreceptor
FT	AT1G65480	Floral integrator
FVE	AT2G19520	Autonomous pathway gene
LD	AT4G02560	Autonomous pathway gene
FLK	AT3G04610	Autonomous pathway gene
FY	AT5G13480	Autonomous pathway gene
FCA	AT4G16280	Autonomous pathway gene
FPA	AT2G43410	Autonomous pathway gene
FLD	AT3G10390	Autonomous pathway gene
FKF1	AT1G68050	Regulator of *CO* expression

Table 1.4 – Genes involved in the floral transition in *Arabidopsis* that will be targeted in lettuce.

Firstly the genes had to exhibit a significant role in regulating flowering in *Arabidopsis*, a number of genes have been identified as being integral to floral induction, some of which affect one or more of the flowering pathways; these genes were prioritised. Secondly genes which when mutated result in delayed/late flowering in *Arabidopsis* were highlighted as potentially useful. Genes involved in the plant's endogenous circadian clock were avoided; genes involved in this process are likely to have pleiotropic effects resulting in undesirable morphological changes to the phenotype of the plant. Finally, the seven genes making up the autonomous pathway were preferentially targeted due to the fact that the effect of mutations within these genes could be overcome with a vernalisation treatment in *Arabidopsis*. Plants with such mutations would be of benefit to both lettuce growers and breeders, see section 1.4.4 for an explanation.

The specific aims of this section were -

(i) To identify homologues of *Arabidopsis* flowering time genes from lettuce. The sequences obtained would cover the full length genomic sequence of the genes targeted, including approximately 1 kb of 5' untranslated region, to include the promoter of the gene.

(ii) To show that the gene sequences obtained are functional orthologues of the *Arabidopsis* flowering time gene by functionally complementing *Arabidopsis* mutants.

1.7.2 Identification of late bolting lettuce lines

The second area of research was concerned with generating lettuce lines displaying a range of bolting phenotypes. Any lines with a desired phenotype would be made available to the industrial partners on this project, Rijk Zwaan®. Due to the current public perception of genetically modified (GM) food, it was important to focus on generating lines with the potential to be used commercially. It was therefore decided, where possible, to work with non-GM plants. The lines identified also needed to produce a robust phenotype and needed to maintain the phenotype of the lettuce cultivar from which they had been generated. This project set out with the objective of obtaining lettuce plants with a delayed bolting phenotype and an increased holding ability in the field. Therefore a number of strategies were undertaken to obtain lettuce lines displaying a wide range of bolting phenotypes.

1.7.2.1 Analysing lines which contain induced mutations

Two lettuce populations which had been mutagenised using EMS, a non-GM technique, were screened for bolting time along with lettuce lines which had been

transformed with the tobacco retrotransposon element, *Tnt1*. LB lines identified will be back-crossed (BC) to WT plants to clean up undesirable background mutations from the mutation(s) causing the LB phenotype, plants would then be self fertilised to create homozygous mutant lines, which could be reassessed for a consistent late bolting phenotype.

1.7.2.2 Identifying natural allelic variation within a wild lettuce diversity set

Natural variation in wild lettuce species was also investigated, naturally occurring allelic variation in genes controlling bolting/flowering time in LB individuals could potentially be breed into cultivated lines, creating new germplasm for the plant breeders at Rijk Zwaan® to work on.

The specific aims of this section were -

(i) To identify LB lettuce lines from commercial lettuce EMS mutagenised populations and the *Tnt1* transformed lines.

(ii) To identify naturally occurring mutations in lettuce target genes that affect bolting time in a diversity set of wild lettuce varieties.

(iii) To create homozygous mutant lines, with a LB phenotype.

1.7.3 Identification of allelic variation in late bolting lettuce lines

The final area of research involved bringing together the work performed in Chapters 3 and 4 in identifying the allelic variation that is causing the phenotypes of the LB lines identified. Two approaches were taken, firstly by looking for variation in the lettuce flowering time gene homologues in the LB lines obtained. Secondly by looking for sequence variation within the full lettuce transcriptome of the LB lines

compared to that of WT using Illumina genome analyser technology. This enables all genetic variation in expressed genes created by the EMS treatment in the LB line to be identified, therefore mutations occurring within non-targeted lettuce flowering time genes and novel lettuce flowering genes can be further analysed for any role in causing the LB phenotype observed. This method has been used with success to discover single nucleotide polymorphisms in *Brassica napus* (Trick *et al*., 2009).

The specific aims of this section were -

(i) To identify mutations within lettuce flowering genes in the LB lettuce lines (whether from the genes targeted, or identified in other genes identified from the transcriptome sequencing).

(ii) To confirm the segregation of the interesting mutation in the line with the LB phenotype throughout the population from which it was generated.

(iii) To investigate whether the expression of any gene containing a mutation which is thought to be causing the LB phenotype is altered during development.

CHAPTER 2

Materials and Methods

Sections 2.1, 2.2 and 2.3 in this chapter outline the standard materials and methods used throughout this project. Detailed methods of specific experiments carried out can be found in the corresponding results sections in chapters 3 - 5.

2.1 Materials

2.1.1 Plant materials

The cultivated lettuce varieties primarily used in this study were *L. sativa* cv. Larissa and *L. sativa* cv. Saladin. Seeds were obtained from Moles Seeds (Colchester, Essex). EMS mutagenised seeds of these lines were generated by and obtained from Dr. Paul Hand (Warwick HRI).

The lettuce wild diversity set screened was made up of 120 lines of *L. saligna*, *L. serriola* and *L. virosa*. The collection of seed was acquired from a number of sources. 50 lines were sourced from the Warwick HRI Genetic Resources Unit, a further 66 lines were provided by Dr. Maria Truco (UC Davis), and 4 lines were provided by Centre for Genetic Resources (CGN), (The Netherlands). Table 1.3 details the breakdown of the lines screened and their source.

T1 and T2 lettuce seeds transformed with the Tobacco *Tnt1* retrotransposon element were provided by Dr. Marianne Mazier (Centre de Recherche Agronomique d'Avignon, France), see section 1.6.3.2 for details. These lines are in the cultivated lettuce variety *L. sativa* cv. Jessy.

Arabidopsis ecotypes Landsberg *erecta* (L*er*) and Columbia (Col-0) seed, along with *ft-1* (L*er* background) mutant seed were provided by Dr. Andrea Massiah (Warwick

HRI). *flk-3* and *flk-4* (Col-0 background) mutant seed was obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). *fld-3* (Col-0 background) mutant seed was provided by Professor Richard Amasino (University of Wisconsin, USA).

2.1.2 General growth conditions

2.1.2.1 Lettuce

All seeds were sown in Levington F2 compost, in modular trays, at a depth of approximately 0.5 cm and were covered in vermiculite, before being watered and stratified in the dark at 4 °C for 48-72 h to achieve uniform germination. When plants had 3-4 true leaves they were transplanted into Levington M2 compost in 5 inch pots for glasshouse experiments or directly into soil for field trials. Plants were grown at Wellesbourne latitude 52°12', unless otherwise stated.

2.1.2.2 *Arabidopsis*

Arabidopsis plants were grown in Levingtons F2 compost containing 6 parts compost, 1 part sand and 1 part vermiculite. Seeds were stratified in the same way as lettuce. Plants were then transferred to growth chambers (Sanyo, MLR-351H) and grown at 22 °C in either SD or LD photoperiods; SDs consisted of 8 h white light (100 $\mu mols^{-1}m^{-2}$), followed by 16 h of dark. LDs consisted of 16 h of white light (100 $\mu mols^{-1}m^{-2}$) followed by 8 h of dark.

2.1.2.3 Glasshouse conditions

Lettuce plants were subjected to glasshouse conditions of 16 h of light, 8 h dark, which was provided by high pressure sodium lamps (400W son-t bulbs), when natural light levels went below 300 W/m^2. Constant day/night temperatures of

18°C/15 °C were maintained. Plants were watered when required and water supplemented with Vitax 2:1:4 (NPK) was applied twice a week. Basal leaves were removed from the plants when required to prevent disease spreading, figure 2.1 shows the glasshouse setup.

Figure 2.1 – Example of the layout of plants growing in glasshouse

Vernalisation of lettuce entailed seedlings with 3-4 true leaves being transferred to 4 °C for 4 weeks, artificial lighting was supplied for 8 h/day to mirror winter conditions.

Alpha block randomisation methods developed by Patterson *et al*., 1978, were applied to plants grown in glasshouse experiments to ensure no individual lines gained preferential growing conditions.

Leaf material was collected from newly expanding leaves and was placed into 1.5 ml microfuge tubes in liquid nitrogen at time points deemed appropriate for individual experiments. The material was stored at -80 °C until required.

2.1.3 Other materials

Lettuce EST and BAC DNA clones were originally generated by the Compositae Genome Project (UC Davis). These clones were made available through the Arizona Genomics Institute (AGI, University of Arizona). The lettuce BAC DNA library was provided by Professor Richard Michelmore (UC Davis).

2.2 General Methods

2.2.1 Lettuce genomic DNA extraction

Genomic DNA was extracted from lettuce leaves using a modified CTAB method. (Doyle and Doyle 1987; Cullings 1992; Porebski *et al*., 1997). Frozen leaf tissue was homogenised using a Dremel drill with an attachment designed to fit a 1.5 ml microfuge tube. Before the sample thawed, 400 µl of CTAB Buffer B (100 mM Tris-HCl, pH 8, 1.4 M NaCl, 20 mM EDTA, 2 % CTAB), was added and the sample was homogenised for a further 20 s. Samples were then incubated at 65 °C for 30 m, before being centrifuged at 13,000 x g in a microfuge for 5 m. The supernatant was then extracted twice with 300 µl chloroform/isoamylalcohol (24:1). 300 µl CTAB Buffer C (1 % CTAB, 10 mM EDTA, 50 mM Tris pH 8) was added to each sample and the tubes were incubated at room temperature for 1 h. The samples were centrifuged at 13,000 x g for 10 m and the pellet was dissolved in 400 µl 1 M CsCl. 800 µl of 100 % ethanol was added, before centrifuging the samples at 13,000 x g for 10 m, the supernatant was removed and two rounds of 70 % ethanol washes were completed before leaving the pellet to air day. Pellets were resuspended in 25 µl TE and were treated with 10 µg/ml RNAse A (Qiagen, Cat. No. 19101).

2.2.2 Rapid Arabidopsis genomic DNA extraction

A modified version of the protocol described by Edwards *et al*., 1991 was used to extract genomic DNA from *Arabidopsis.*

Leaf tissue was homogenised using a Dremel drill with an attachment designed to fit a 1.5 ml microfuge tube. 400 µl of Extraction Buffer (200 mM Tris HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5 % SDS), was added to each sample and was further homogenised. The samples were centrifuged for 2 m at 13,000 x g, supernatant was

placed into a new microfuge tube before extracting with chloroform:isoamyl alcohol (24:1). The samples were then precipitated with 3 M sodium acetate pH 5.2 and isopropanol at -20 °C for 2 h. Genomic DNA was collected by centrifugation before being resuspended in 50 µl TE and treated with 10 µg/ml RNAse A.

2.2.3 PCR

Standard PCR was performed in a volume of 20 µl. Each reaction contained 0.4 U KOD Hot Start DNA Polymerase (Merck Chemicals, Cat. No. 71086), forward and reverse primers were used at a concentration of 0.3 µM, 0.2 mM each dNTP, 2 mM $MgSO_4$ and 1 µl of template at 100 ng/µl. PCR reactions were carried out with an initial denaturation step of 95 °C for 2 m followed by cycles of denaturation at 95 °C for 20 s, annealing at a temperature deemed appropriate for the primer pair for 10 s and extension at 70 °C for 20 s/ kb of product.

PCR products of >3 kb, which could not be successfully amplified using KOD Hot Start DNA Polymerase, were amplified using KOD Xtreme™ Hot Start DNA Polymerase (Merck Chemicals, Cat. No. 71975). Reactions were carried out in a 50 µl volume using 1 U of Polymerase; primer and dNTP concentrations were as used with standard PCR. An initial denaturation step of 95 °C for 2 m was followed by 35 cycles of 98 °C for 10 s, an annealing temperature deemed suitable for the primer pair used for 30 s, followed by an extension at 68 °C for 25 s/kb of product.

All PCR primers were designed using the internet package, Primer 3 (Rozen and Skaletsky, 2000) (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primers were designed to be approximately 25 bp in length and to have a Tm of 60-65 °C. Primers were synthesised by Invitrogen Ltd.

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to visualise PCR products amplified. 5 µl Orange G (Sigma Aldrich, Cat. No. O3756) loading dye was added to each PCR product before running on 0.8-2 % agarose (Invitrogen Ltd., Cat. No. 15510) / 1x TAE buffer (VWR International, Cat. No. 44125D) gels containing Ethidium bromide, depending on the size of the expected product. Gels were run in electrophoresis tanks containing 1x TAE buffer at a voltage and for a period dependent on the size of gel and products expected. 3 µl of 1 kb Plus DNA ladder (Invitrogen Ltd., Cat. No. 10787) was loaded onto each gel to use as a guide for estimating the size of the PCR product amplified. Images of the gels were photographed using a Syngene G:BOX gel documentation system.

2.2.5 Purification of PCR products

PCR products were purified by either excising the band of interest from an agarose gel under ultra-violet light (BioRad UV Transilluminator 2000), and using the QIAquick Gel Extraction Kit (Qiagen, Cat. No. 28704), or by purifying the PCR product directly from the PCR reaction using the QIAquick PCR Purification Kit (Qiagen, Cat. No. 28104). Occasionally products of low concentration were purified using the MinElute PCR Purification Kit (Qiagen, Cat. No. 28004). All PCR products were purified following the manufacturer's protocol. Each sample was eluted from the spin column with 30 µl sterile distilled water, MinElute samples were eluted in 10 µl.

2.2.6 RNA extractions and cDNA synthesis

Total RNA was extracted using the Z6 Buffer method (Logemann *et al*., 1987). Frozen leaf tissue was homogenised using a Dremel drill with an attachment designed to fit a 1.5 ml microfuge tube, 400µl buffer was added and the tissue was homogenised until completely ground. 400 µl Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added and the samples were mixed prior to being centrifuged for 15 m at 13,000 x g. The supernatant was removed into a sterile microfuge tube and was precipitated with 1 M Acetic acid and 100 % Ethanol at -20 °C for 1 h. A pellet was collected by centrifuging at 13,000 x g for 10 m and was dissolved as much as possible in 500 µl 3 M Sodium Acetate pH 5.2. The samples were washed twice with 70 % ethanol, before resuspending in 50 µl of sterile distilled water at 65 °C.

5 µg of total RNA was treated with Turbo DNA-free™ DNase (Ambion Inc, Cat. No. AM1907), following the manufacturer's guidelines. First round cDNA was synthesised using Thermoscript reverse transcriptase (Invitrogen Ltd, Cat. No. 11146024) following the manufacturer's guidelines, oligo(dT) was used as the primer in the synthesis reaction.

cDNA and RNA samples were quantified using the NanoDrop™ ND-1000 Spectrophotometer (Thermo Scientific). A volume of 1.2 µl was quantified.

2.2.7 Genome walking

This protocol is based on the method described by Clontech Ltd. 2 µg of genomic DNA or 700 ng of BAC DNA was digested with five individual blunt-ended restriction enzymes; *DraI, EcoRV, PvuII, ScaI* and *SmaI,* to create five libraries. Two adaptor primers; 5'-CTA ATA CGA CTC ACT ATA GGG CTC GAG CGG CCG CCC GGG CAG GT-3' and PO_4-ACC TGC CC-NH_2 were annealed to each

other, at a final concentration of 50 µM by heating to 95 °C and allowing the mix to return, slowly to room temperature. The adaptors were then ligated to the ends of the digested DNA using 5 U T4 DNA Ligase (Roche Diagnostics, Cat. No. 10799009001), overnight at 16 °C.

Two primers were designed to anneal to the adaptor sequences (AP1 and AP2), and two were designed to the known DNA sequence from which the genome walking was to take place (GSP1 and GSP2). Primers were optimally designed to be 27 bp in length, with a Tm of 72 °C and a GC-content of 40-60 %. A first round touch-down PCR was performed using 0.5 µl of the ligation as template in a 10 µl reaction volume, the primers GSP1 and AP1 were used at a concentration of 0.4 µM, other reagents used in the reaction were at concentrations stated in section 2.2.3. The PCR cycling conditions were as follows, an initial denaturation step at 95 °C of 2 m, was followed by six cycles of touch down; 95 °C for 20 s, 72 °C for 10 s, with a decrement of 1 °C per cycle and 70 °C for 1 m 30 s. A further 24 cycles of 95 °C for 20 s, 67 °C for 10 s and 70 °C for 1m 30 s was followed by a final extension step of 70 °C for 4 m. A second round of touch-down PCR was performed, using AP2 and GSP2 primers with 1 µl of first round PCR product as template. Products were run on a 1 % gel and were sequence verified before being cloned, see section 2.2.8 for details.

2.2.8 Molecular cloning

Unless otherwise stated, cloning was carried out using the pGEM®-t Easy Vector System (Promega Ltd., Cat. No. A1360), see figure 2.2 for details, following the manufacturer's guidelines. Ligations were set up using a 3:1 insert: vector ratio, the amount of insert to add was calculated using the equation below.

$$\frac{\text{(vector ng)(insert kb)}}{\text{(vector)}} \quad x \quad \text{(molar ratio insert/vector)} = \text{insert ng}$$

Transformations were performed by electroporation into EC100 electrocompetent cells (Cambio Ltd., Cat. No. EC10005). Cells were plated onto LB/agar containing the appropriate antibiotics, for cloning using the pGEM®-t Easy Vector System and the pEZSeq™ vector ampicillin (Melford Laboratories Ltd., Cat. No. A0104) was used at 100 µg/ml. Kanamycin (Melford Laboratories Ltd., Cat. No. K0126) at 50 µg/ml was used with the pBluescript II SK(+) vector.

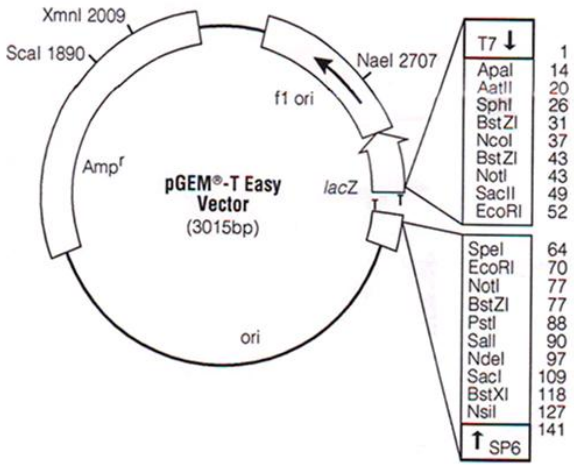


Figure 2.2 – pGEM®-t Easy Vector

2.2.9 Plasmid DNA extraction

5 ml LB media (VWR International, Cat. No. 1.10285), containing the appropriate antibiotic were inoculated with individual bacterial colonies and were grown overnight at 220 x g and 37 °C. Bacterial cells were harvested by centrifugation at 3000 x g for 10 m. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen, Cat. No. 27106), following the manufacturer's guidelines. DNA was eluted from the spin columns using 50 µl of sterile distilled water.

2.2.10 Sequencing

Samples were sequenced using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Cat. No. 4337456). Each sequencing reaction contained an appropriate concentration of template, according to the manufacturer's guidelines, along with 2 µl Big Dye and 4.2 µM primer in a final volume of 10 µl. Cycling conditions comprised of 1 cycle of 96 °C for 1 m followed by 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 4 m. Sequence reactions were run on an ABI 3730 sequencing machine in the Warwick HRI Genomics Resource Centre.

Sequence files were analysed using EditSeq and MegAlign software packages of DNAStar (DNAStar Inc.). Chromatograms were viewed using Chromas 2.11 (Technelysium Pty Ltd).

Phylogenetic trees were constructed as follows; sequences were collated in a Notepad document. The sequences were aligned using Clustal X (Thompson *et al*., 1997), and were edited using Jalview (Clamp *et al*., 2004). Neighbour-joining trees were constructed and bootstrap values were calculated using 1000 replicates using Clustal X. The trees were viewed using NJPlot (Perrière and Gouy, 1996).

2.2.11 Real time-PCR

Quantitative real-time PCR was carried out using an I-Cycler (Bio-Rad Laboratories, iCycler Thermal Cycler). Each reaction contained 1 µl of cDNA, 1x PCR Mastermix containing SYBR green (Eurogentec Ltd., Cat. No. RT-SY2X-03+WOUFL) and 10 µM of each primer. An initial step of 55 °C for 2 m was carried out to activate Uracil-N-glycosylase. This was followed by 50 cycles of 95 °C denaturation, followed by annealing at 61 °C for 45 s, before a cycle of 95 °C for 1 m and 55 °C for 1 m. A melt curve was produced from 80 cycles of 55 °C for 10 s, increasing the

temperature by 0.5 °C per cycle. Primers used in these experiments were designed using Primer 3, the Eurofins MWG Operon oligo analysis tool (http://www.operon.com/technical/toolkit.aspx) and the mfold oligo analysis package (http://www.bioinfo.rpi.edu/applications/mfold/cgi-bin/dna-form1.cgi). Each sample was run in triplicate and an average value based on the number of PCR cycles when a product is first detected was calculated. The data obtained was normalized to the expression of the house-keeping gene *ELONGATION FACTOR 1 Alpha (EF1α)*. A normalised figure was obtained by dividing the expression profile of *EF1α* at the same time point. Standard curves, using serial dilutions were plotted using cDNA synthesised from leaf material harvested at various time-points in a 24 h period.

2.2.12 Southern blotting

Approximately 10 µg lettuce genomic DNA was digested using 15-30 U of restriction enzyme. Reactions took place overnight at 37 °C in 30 µl volumes. Digested products were run on 0.8 % agarose gels and were Southern blotted (Southern 1975), overnight on to a positively charged nylon membrane (Roche Diagnostics, Cat. No. 11417240001).

2.2.13 Preparation of radioactive probes and hybridisation reactions

25 ng of purified DNA was diluted 1:10 with sterile distilled water and was denatured at 95 °C for 5 m before snap cooling on ice. The DNA was added to a tube containing the Rediprime™ II DNA Labeling System (GE Healthcare, Cat. No. RPN 1633) mix. 5 µl of ^{32}P dCTP (Perkin-Elmer, Cat. No. NEG513H), was added to each tube and the reaction was incubated at 37 °C for 1 h, before being purified on an Illustra ProbeQuant™ G-50 micro column (GE Healthcare, Cat. No. 28-9034-08).

Membranes were pre-hybridised in 30 ml of Hybridisation buffer, (0.5 M Sodium Phosphate buffer, pH 7.2, 7 % SDS, 1 mM EDTA, 10 µg/ml sheared denatured salmon sperm DNA), at 65 °C for 4 h. 15 µl of probe which had been denatured at 95 °C for 5 m was then added to each membrane. The hybridisation took place overnight at 65 °C.

The membranes were washed several times at 65 °C, using 2x SSC, 0.1 % SDS – 0.1x SSC,0.1 % SDS for 20 m – 1 h. When radioactive counts had been reduced to ~5 per second, the membranes were wrapped in cling-film and were placed in phosphoimager cassettes overnight before exposing using the Phosphoimager SI system (Molecular Dynamics).

2.2.14 Colony lifts

Colonies were transferred from agarose plates to positively charged nylon membranes as follows; filters were placed onto the plates for 15min before being placed face up in 1 ml 10 % SDS Buffer for 5 m. The filter was then transferred to 1 ml denaturation buffer (1.5 M NaCl; 0.5 M NaOH), for 3 m before being placed in neutralisation buffer (0.5 M Tris; 1.5 M NaCl pH 7), for a further 3 m, this final step was repeated. Each filter was then washed with 3x SSC Buffer for 10 m. The DNA was fixed to the filters by baking at 120 °C for 2 h.

2.2.15 TILLING (performed by Genomics Centre, John Innes Centre, Norwich)

Labeled primers; the labels were 5' Fam on the forward primers and 5' Hex on the reverse primers, and non-labeled primers were synthesised by Sigma-Aldrich® and were mixed together in a 3:2 (labeled:unlabeled) ratio; PCR amplification of target genomic DNA sequences was performed using 0.2 µM forward and reverse

unlabeled/labeled primers, for primer details see Appendix II, 0.5 U Takara ExTaq™ (Takara Bio. Inc., Cat. No. RR001A), and 0.8 mM dNTPS. The reactions were cycled as follows, an initial denaturation step of 95 °C for 2 m was followed by 8 cycles of 94 °C for 20 s, 63 °C for 30 cycles, with a decrement of 1 °C per cycle, 72 °C for 1 m. A further 45 cycles of 94 °C for 20 s, 55 °C for 30 s, 72 °C for 1 m with a final extension of 72 °C for 5 m was performed before the amplified products were denatured at 99 °C for 10 m and slowly reannealed back together using 70 cycles of 70 °C for 20 s with a decrement of 0.3 °C per cycle. The reannealed PCR products will contain products containing mismatches where SNPs are present in sequences, see figure 1.9. The enzyme CEL1 recognises such mismatches and cleaves the products, 1 µl of CEL1 (Home-made, JIC), was added to 5 µl of PCR product and is incubated for 15 m at 45 °C. The reaction was stopped by adding 5 µl of 150 mM EDTA to the samples and was precipitated using isopropanol. The air dried samples had 10 µl of Rox 1 kb ladder (Web Scientific Ltd. Cat. No. MRK1000) : Hi-Di™ formamide (Applied Biosystems™ Cat. No. 4440753), (1:100) mix added to them and were denatured at 95 °C for 3 m. The samples were run on the ABI3730XL (Applied Biosystems™) and the data produced was analysed using Genemapper 4.0 (Applied Biosystems™). SNPs identified were confirmed by PCR using gene specific primers followed by sequencing or digesting the amplified products obtained, see sections 5.2.13 and 5.2.1.4 for details.

2.2.16 Backcrossing

Lettuce cultivars are highly inbred and so need to be manipulated to enable crossing, see figure 2.3 for details. Petals from open lettuce flowers were trimmed before pollen was washed off the emerging stigmas by a directed stream of water. The

stigmas were then dried with a stream of air before the foreign pollen was applied (Oliver 1910; Ryder and Johnson 1974).

Backcrossing took place before midday each morning. WT flowers were used as the pollen donor; the flowers from these plants tend to open earlier and produce more pollen than the mutant lines.

Figure 2.3 – Method used to back-cross lettuce plants. Petals are trimmed (A), before a stream of water is applied using a syringe to remove any self pollen (B). The stigmas are dried with a stream of air (C), before the pollen from the donor plant was applied (D)

2.3 Statistical analysis of data

Where appropriate, bolting data was analysed using one-way analysis of variance (ANOVA) tests, which were carried out to assess the significance of any differences in the bolting time of WT plants compared with lines bolting later which were identified in the lettuce populations screened. Unless otherwise stated, least significance values (l.s.d.; α=0.05), were used from the ANOVA output to compare the significance of the average number of days to bolt recorded for WT plants with the other lines screened. ANOVA tests were also used to look for significant

differences in flowering time in *Arabidopsis* complementation experiments. The output data from ANOVA calculations can be found in Appendix VIII.

Quantile (Q-Q) plots (Wilk and Gnanadesikan, 1968) were also generated where appropriate to assess the variation in bolting time for the Larissa and Saladin EMS mutagenised lettuce populations. This analysis compared the average days to bolt for each line with a Gaussian (normal) distribution calculated from the average days to bolt for the population as a whole. The output from this analysis was two-fold; firstly it showed whether the data fitted a Gaussian distribution at a 90 % confidence level. If this was the case, the data can be used in a second way; to look at the probability that a line will bolt later than the rest of the population, this was calculated by identifying the normalised score of the interesting lines plotted on the y-axis and using cumulative probabilities of the standard normal distribution statistical tables (Neter *et al*., 1996).

All statistical calculations were carried out using the Genstat for Windows version 12 (VSN International, Herts., UK).

Where the average number of days to bolt per line has been used, standard errors of the mean (SEM) have been calculated and quoted in the text where appropriate, these values have also been plotted on graphs throughout this thesis, unless otherwise stated.

CHAPTER 3

Isolation of flowering time gene homologues from lettuce

3.1 Introduction

This chapter discusses the results obtained from isolating full length sequences of the genes involved in the regulation of flowering time from lettuce.

Initially the lettuce EST database, CGP1, was used to find ESTs with homology to genes known to be involved in regulating flowering time in *Arabidopsis*. High levels of homology between *Arabidopsis* and lettuce ESTs were observed for *FVE, CRY2, CO,* and *FLK*. PCR primers were designed based on the lettuce ESTs with the aim of amplifying fragments which could then be used to probe a genomic or cDNA library. Genomic/cDNA fragments representing regions of lettuce gene homologues of *CO*, *CRY2* and *FVE* were obtained.

The subsequent release of the CGP2 database and the availability, through the Arizona Genome Institute (AGI) of lettuce EST clones in February 2007, along with availability of a lettuce BAC DNA library through the collaboration set up with Professor Michelmore (UC Davis), meant that more rapid progress could be made.

Instead of using PCR to generate probe fragments, the lettuce EST clones showing the highest homology to the selected *Arabidopsis* flowering time genes were ordered from AGI and these were used as probes to screen the lettuce BAC DNA library. Data was obtained for 12 target genes, however the list of target genes was rationalised at a later date and reduced to 8 genes (*FT, FLK, FLD, LD, FVE, FY, FCA* and *FPA*) which were prioritised in this project.

This chapter details the genes targeted and describes the results obtained from isolating them in lettuce. The details of all the PCR and genome walk primers mentioned in this chapter can be found in Appendix II.

3.2 Materials and Methods

3.2.1 Screening the CGP lettuce EST database

The database was screened to identify EST sequences showing similarity to *Arabidopsis* flowering time genes that were targeted in this project. CPG1 is made up of sequences generated from material collected from *L. sativa* cv. Salinas; whereas CPG2 incorporates sequences from *L. serriola, L. saligna, L. virosa* and *L. perennis* as well as *L. sativa* cv. Salinas, see Appendix III, table A2, for details of materials used to construct EST sequences. Both databases were comprised of sequences represented by single ESTs, covering around ~700-900 bp and individual contigs, which were created by piecing together ESTs. Some contigs covered entire gene sequences. The *Arabidopsis* nucleotide and amino acid sequences were BLAST searched against the EST database.

The CGP1 database can be found at http://cgpdb.ucdavis.edu/sitemap.html, the BLAST search against Lettuce Assembly option was chosen along with the blastn option for nucleotide sequences and the tblastn option for amino acid sequences. The hits obtained, were either in the form of individual ESTs or contigs created by piecing together more than one EST showing homology to an *Arabidopsis* gene. The EST sequence information could be retrieved by using the EST Retrieval from Database option, contig information was viewed by selecting the Contig Assembly Viewer option. The CGP2 database can be found at http://cgpdb.ucdavis.edu/cgpdb

2/, the BLAST search against EST assemblies option was chosen, and the lettuce 5X all seqs 41K Set was selected, this meant the *Arabidopsis* sequences were BLASTed against EST information from all five lettuce species represented within CGP2. Again blastn and tblastn BLAST programs were chosen dependent on whether the sequence was comprised of nucleotides or amino acids.

EST clones displaying the highest level of homology were identified by aligning sequences and looking at percentage identity, and through creating phylogenetic trees using DNASTAR Lasergene 8 software and Clustal X as described in section 2.2.10. The EST clones of interest were ordered from AGI.

3.2.2 Verification and sequencing of EST clones

The lettuce ESTs are cloned into the 3.7 kb pBRcDNA-Sfi-AB vector (figure 3.1) and were obtained from AGI as bacterial cultures. Clones were grown on LB/agar plates containing 100 µg/µl ampicillin (Melford Laboratories Ltd, Cat. No. A0104). Plasmid DNA was isolated and M13 forward, 5' GTA AAA CGA CGG CCA G 3', and M13 reverse primers, 5' CAG GAA ACA GCT ATG AC 3', were used to sequence each clone, as described in sections 2.2.9 and 2.2.10. EST specific primers were also designed to obtain as much EST sequence as possible, see Appendix II for primer details.

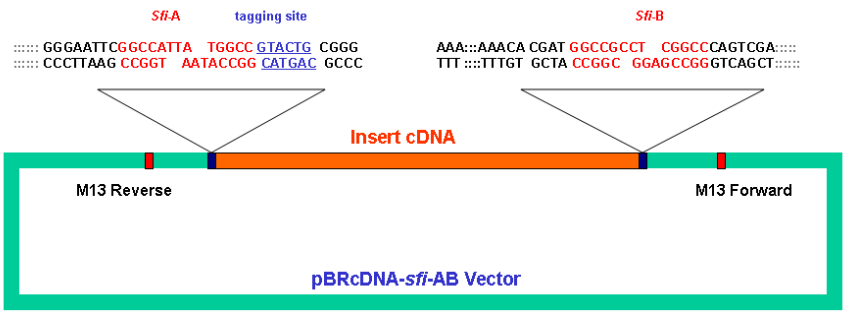


Figure 3.1 – pBRcDNA-Sfi-AB Vector. EST insert is digested from vector by using *SfiI*. Figure taken from - cgpdb.ucdavis.edu/Library_Construction/

3.2.3 Screening a lettuce BAC DNA Library to identify putative lettuce flowering time gene homologues

EST clones with similarity to *Arabidopsis* flowering time genes were used to screen a *L. sativa* cv. Diana BAC DNA library. The BAC library is comprised of approximately 50,000 clones and is spotted on five membranes. Each EST insert was digested from the pBRcDNA-Sfi-AB plasmid DNA using 15U of *SfiI* at 50 °C for 1 h. The insert was excised from an agarose gel and was purified as described in section 2.25.

EST Plasmid DNA was labeled and BAC library membranes were hybridised and treated as described in section 2.2.13.

Positive hits identified on the membranes were located as follows; each membrane is divided into six fields, see figure 3.2, each field contains 384 squares; these squares represent the identification of each BAC. Within each square there are 16 positions where eight clones are spotted in duplicate. The pattern of the spotted clones in conjunction with reference tables generates the plate address of the BAC clone of interest. BAC clones of interest were ordered from AGI, the details of which can be found in Appendix IV, table A3.

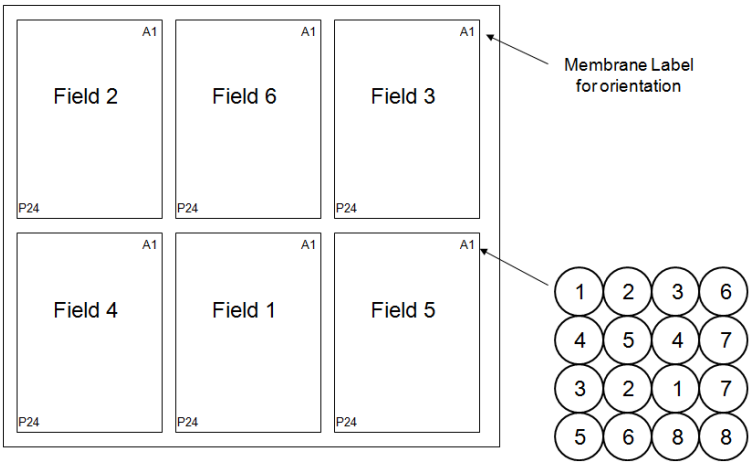


Figure 3.2 – An example of a BAC library membrane. Each membrane contains information from 48 x 384 well plates. Within each well position are 16 spots representing eight clones spotted in duplicate.

3.2.4 BAC clone verification and DNA isolation

To ensure the BAC clones contained the sequences of interest a set of primers were designed to the sequence of each EST that had been used to probe the library. The primers were designed to a single exon within each EST sequence; this was determined by lining up *Arabidopsis* genomic and coding sequences with the lettuce EST sequence. This ensured that an exonic product of a known size could be generated. Each product was sequenced to ensure that the expected gene sequence had been amplified.

Each BAC clone insert is approximately 110 kb in size, and was supplied as a bacterial culture in the 9.3 kb pECSBAC4 vector which contains a chloramphenicol resistance gene. Following manufacturer's guidelines, the Large Construct Kit (Qiagen, Cat. No. 12462), was initially used to extract BAC DNA from a 500 ml overnight culture. An alternative method was developed which produced a higher concentration of BAC DNA; 200 ml of overnight culture was grown, the cells were harvested and were treated with 5 ml of Qiagen buffers, P1, P2 and N3. The supernatant was then precipitated with isopropanol, washed with 70 % ethanol and resuspended in 500 µl TE. The sample was then extracted with chloroform:isoamyl alcohol, (24:1), precipitated with ethanol, washed in 70 % ethanol and re-suspended in 100ul TE.

3.2.5 Sub-cloning of BAC DNA fragments containing lettuce target genes

Approximately 3 µg of BAC DNA was digested with restriction enzymes that did not cut within the EST sequence and was Southern blotted overnight, as described in section 2.2.12. The membranes were probed using radioactively labeled ESTs as described in section 2.2.13 to identify the fragments to be sub-cloned. Two cloning

methods were used to sub-clone BAC DNA fragments. The first method was as follows; the fragment of interest was excised from the agarose gel and was gel purified, see section 2.2.5 for details. pBluescript II SK (+) (Agilent Technologies UK Ltd., Cat. No. 212207) plasmid DNA was digested using 10 U of the restriction enzyme used to digest the BAC DNA. The products were ligated together and transformed into EC100 electrocompetent cells and selected on 50 µg/ml Kanamycin. Positive colonies were PCR screened for inserts before plasmid DNA was extracted and sequenced as described in sections 2.2.9 and 2.2.10.

The second method involved shearing the BAC DNA into fragments of ~2.5 kb, which were then sub-cloned, colonies which grew were transferred to a positively charged nylon membrane and were probed with the radioactively labeled EST to see which of the colonies contained the gene of interest. Approximately 50 µg of BAC DNA was added to 750 µl nebulisation buffer (4 ml TE pH 8 and 1 ml 100 % Glycerol), and was mixed before being transferred to a nebulisation unit (Invitrogen Ltd, Cat. No. K7025-05). Compressed air was applied to the nebulisation unit at 10 psi for 2 m. The sheared DNA was precipitated with 80 µl 3 M Sodium acetate pH 5.2, 700 µl 100 % isopropanol and 4 µl 1 mg/ml glycogen (Fermentas Ltd, Cat. No. R0561). The precipitated product was resuspended in 50 µl sterile distilled water and was purified using the PureLink™ PCR Purification Kit (Invitrogen, K3100-01). The purified product was end-repaired using the End-It™ DNA End-Repair Kit (Epicentre® Biotechnologies, Cat. No. ER0720), following the manufacturer's guidelines. The product was run on a 0.5 % agarose gel and products of ~2 kb were excised and purified using the peqGOLD Microspin Gel Extraction Kit (PeqLab Biotechnologie GmbH, Cat. No. 12-6294-01), following the manufacturer's guidelines.

The eluted product was cloned into the pEZseq™ vector (Lucigen® Corporation, Cat. No. 40464-2), following the manufacturer's guidelines, see figure 3.3.

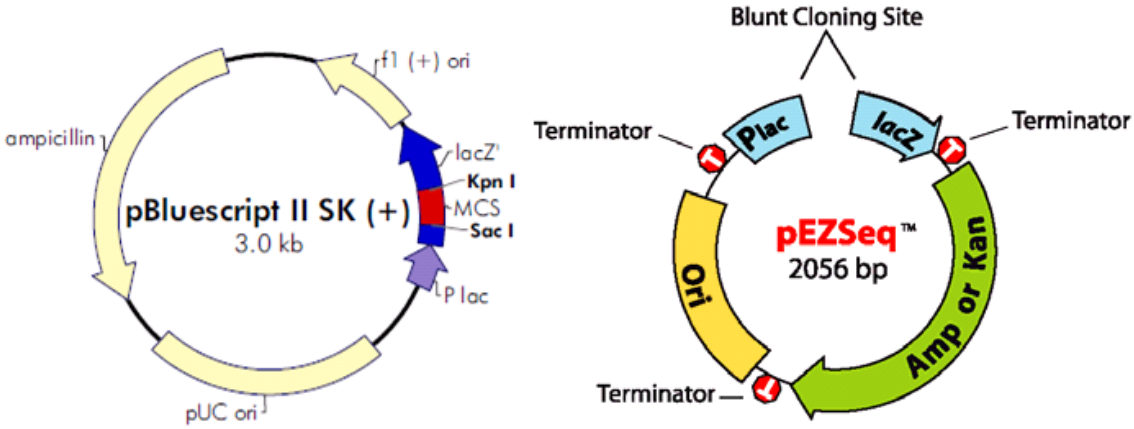


Figure 3.3 – Map of vectors pBluescript II SK(+) and pEZSeq™ used in sub-cloning experiments

A colony lift was then performed on colonies which grew, as described in section 2.2.14. The appropriate EST was radioactively labelled and was used to probe the filters, as described in section 2.2.13.

Where sub-cloning of fragments containing the gene of interest proved difficult, Genome Walking from BAC DNA was employed. Gene specific primers were designed to the EST sequence used to obtain the BAC clone, they were used to walk out of the fragment into new sequence. For details of the protocol used see section 2.2.7.

Other gene sequences were obtained by cloning PCR products. In cases where EST sequences covered a large region of the coding sequence, PCR primers were designed to the extreme 3' and 5' ends of the EST sequence and PCR amplification using KOD Taq Polymerase (see section 2.2.3), was used to obtain large amounts of sequence information; especially exon/intron boundaries. These products were cloned and sequenced as described in sections 2.2.8 and 2.2.10.

3.2.6 Transformation of putative lettuce flowering time genes into Arabidopsis

Full-length cDNA coding sequence of three lettuce gene homologues; *FT*, *FLK* and *FLD* were cloned into the pGEM®-t Easy vector, as described in section 2.2.8. PCR primers with att sites, specific for Invitrogen Gateway® cloning were designed and a 2-step PCR reaction was performed, 4 cycles of 95 °C for 20 s, 55 °C 10 s, 70 °C 1 m were followed by 20 cycles 95 °C for 20 s, 65 °C 10 s, 70 °C 1 m. The PCR reaction mix used is described in section 2.2.3. The products were run on a 1 % gel and were excised and purified, as described in sections 2.2.4 and 2.2.5, before being cloned into the pDONR^{TM}207 vector (Invitrogen Ltd., obtained from Sue Donovan, Warwick HRI), see figure 3.4, using Gateway® BP Clonase® II enzyme mix (Invitrogen, Cat No. 11789-020) following the manufacturer's guidelines. Using Gateway® LR Clonase® II enzyme mix (Invitrogen, Cat. No. 11791-043), the BP transformants were cloned into the pB2GW7 binary vector, (Invitrogen Ltd., obtained from Andrew Taylor, WHRI) see figure 3.4, to produce pB*Ls*FT, pB*Ls*FLK and pB*Ls*FLD vectors, following the manufacturer's guidelines. The *bar* gene in the pB2GW7 vector confers resistance to the herbicide Glufosinate-ammonium, the vector also contains a CaMV 35S promoter, which drives the expression of the gene of interest cloned into the vector. Plasmid DNA was isolated and was sequenced verified using gene-specific primers, see section 2.2.9 and 2.2.10 for details. pB*Ls*FT, pB*Ls*FLK and pB*Ls*FLD were then transformed using electroporation, into the *Agrobacterium* strain c58pGV3101, obtained from Dr. Karl Morris (Warwick HRI) and were plated onto LB/agar containing 100 µg/µl Spectinomycin, 25 µg/µl Gentamicin and 50 µg/µl Rifampicin. The *Agrobacterium* colonies were then screened for the presence of the appropriate gene by PCR using gene-specific primers.

Arabidopsis plants were transformed using the floral dip method (Clough and Bent, 1998). The mutants transformed were *ft-1* (in the L*er* background); *flk-4* and *fld-3* (both in the Col-0 background). 10-20 plants were transformed with each clone, were grown in LD conditions (16 h light, 22 °C) in growth chamber cabinets (Sanyo MLR Plant Growth Chamber) and were allowed to flower and set seed. Seeds were stored at 4 °C. T_1 generation plants were sown in soil drenched with the BASTA herbicide Kaspar^{TM} (Certis), containing the active ingredient, Glufosinate-ammonium at a concentration of 150 mg/L. Seeds were stratified at 4 °C for 72 h before being transferred to LD conditions, the seeds were subjected to four further BASTA treatments. The putative transgenic plants which survived were scored for flowering (in all research journals *Arabidopsis* plants are scored for flowering as opposed to bolting), by recording the number of leaves present when the plant to produce a 1cm bolt. Leaf material was collected from these plants and genomic DNA was extracted, described section 2.2.1, and was screened by PCR using gene-specific primers, described section 2.2.3, to confirm the presence of the transgene. T2 seed was collected from these plants and was screened for flowering time to ensure the phenotype was segregating throughout the population as expected.

This work was carried out using *Ls*FT, *Ls*FLK and *Ls*FLD only. Complementation studies were not carried out for the remaining lettuce genes identified due to the amount of time available for the project, (some full length sequences for the lettuce genes targeted were not obtained until later on in the project). A lack of space and the ability to obtain relevant mutant seed also prevented more complementation work from being carried out.

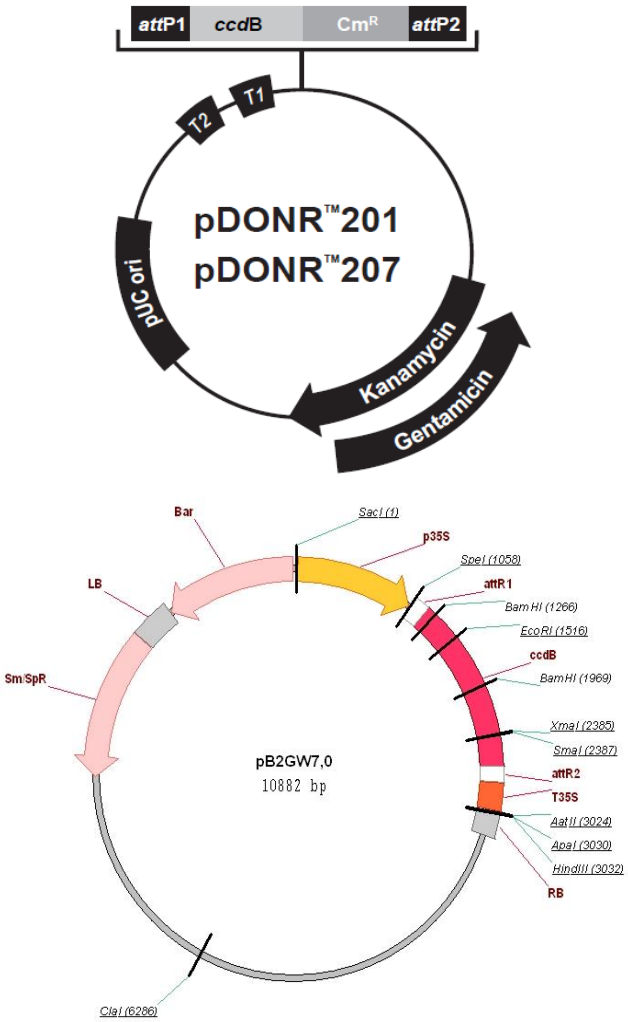


Figure 3.4 – Maps of the vectors used in Gateway® Cloning

3.3 Results and Discussion

3.3.1 EST Database searches

Information from studies in *Arabidopsis* have identified unique or highly conserved domains within each of the flowering time genes targeted. This information was used to help identify which ESTs represented potential lettuce gene homologues.

Lettuce ESTs with homology to all of the genes targeted were identified from the CGP1 and CGP2 databases. *Arabidopsis* target gene sequences, both at the nucleotide level and the amino acid level, were BLASTed against the databases, using the methods explained in section 3.2.1. ESTs and contigs with levels of

identity varying from 20-80 % were obtained. Some *Arabidopsis* gene sequences were represented by a single EST, for example, *FT* and *LD*. These were strong candidates to follow up; other genes were represented by many expressed sequences, for example *FY* and *FCA*. Table 3.1 contains a summary of the hits obtained from the database searches, along with details of the range of percentage identity to the target *Arabidopsis* gene.

Gene	CGP1 Number of Contig/EST hits	CGP2 Number of Contig/EST hits	Identity with *Arabidopsis* at amino acid level (%)
FT	0	1	75
FLK	12	12	24 - 60
FLD	3	3	31 - 65.6
LD	0	1	57.4
FVE	9	9	21 - 83
FY	12	12	23 - 78.5
FCA	12	13	27 - 81.8
FPA	8	2	23 - 40
CO	11	12	34 - 59
CRY2	8	8	54 - 81
FKF1	11	7	28 - 73
FLC	12	12	40 - 57

Table 3.1 – Summary of lettuce ESTs and contigs displaying homology to *Arabidopsis* flowering time genes identified in CGP1 and CGP2 databases

All of the ESTs and contigs in table 3.1 were analysed in more detail to identify the most appropriate sequence(s) on which to base future experiments. The following sections discuss this work in more detail.

3.3.2 FT

3.3.2.1 Introduction

FT promotes flowering in *Arabidopsis*. *AtFT* is responsive to signals from several flowering pathways and as such is described as a floral integrator, see figure 1.2. *AtFT* encodes a protein similar to phosphatidylethanolamine-binding protein (PEBP) (Kardailsky *et al*., 1999; Kobayashi *et al*., 1997), also referred to as a Raf kinase

inhibitor protein (RKIP). The RKIP domain is 472 bp in length and covers amino acid residues 8-165. There are six RKIP family proteins in *Arabidopsis*; one of which is *TWIN SISTER OF FT (TSF),* which shows 81.8 % identity with *AtFT* and has been shown to act redundantly with *AtFT* (Yamaguchi *et al*., 2005). *TERMINAL FLOWER 1* (*TFL1*) is another, which is a floral repressor and acts in an antagonistic manner to *AtFT* (Shannon and Meeks-Wagner 1991), it also has a role in the maintenance of inflorescence meristem identity (Bradley *et al*., 1997). A third RKIP protein, *MOTHER OF FT (MFT),* may also promote flowering like *AtFT* as research has shown that constitutive expression of this gene causes earlier flowering under LD conditions (Yoo *et al*., 2004).

Floral transition by the PEBP/RKIP family is conserved in other plants including SD plants such as rice and other LD plants, for example *Pisum*. *HEADING DATE 3a (Hd3a),* is the *FT* orthologue in rice and plays an important role in promoting flowering in SD conditions (Kojima *et al*., 2002). *Hd3a* homologues in rice have been identified which have also been shown to be implicated in the regulation of flowering (Izawa *et al*., 2002, Kojima, 2002). The activity of *LATE FLOWERING (LF),* a *TFL1* homologue in *Pisum* causes a delay in the floral transition (Foucher *et al*., 2003).

It is the *At*FT/Hd3a protein which appears to act as the long distance signal that induces flowering in *Arabidopsis* and rice (Corbesier *et al*., 2007; Jaeger and Wigge, 2007; Mathieu *et al*., 2007; Tamaki *et al*., 2007).

3.3.2.2 Isolation of a lettuce *FT* (*LsFT*)

Only one EST displayed any significant homology to the *Arabidopsis* FT protein; EST CLSS2320.b1_O03.ab1, which covered the whole of the 175 amino acid coding

sequence of *At*FT, figure 3.5. This was unexpected as *AtFT* is part of a large gene family in *Arabidopsis*, and is highly conserved with *TSF*, also more than one *FT* homologue exists in rice (Izawa *et al*., 2002, Kojima *et al*., 2002), and sunflower which, like lettuce is a member of the *Compositae* family (Blackman *et al*., 2010). Blackman *et al*., (2010), identified four sunflower *FT* homologues all of which are highly conserved with the lettuce gene homologue described in this section. EST CLSS2320.b1_O03.ab1 contains the two conserved FT amino acid residues at position 85 (tyrosine) and position 140 (glutamine), that functionally distinguish FT orthologues from TFL1 (Hanzawa *et al*., 2005 and Ahn *et al*., 2006).

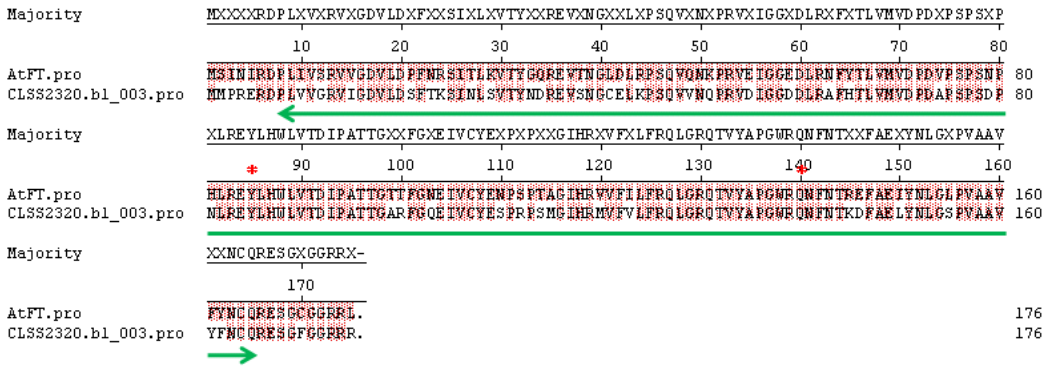


Figure 3.5 – Amino acid sequence comparison of *Arabidopsis* FT with lettuce EST CLSS2320.b1_O03.ab1. The RKIP domain is underlined with a green arrow. The two amino acid residues that functionally distinguish FT from TFL1 are highlighted with a red asterisk.

EST CLSS2320.b1_O03.ab1, was expressed in tissue collected from *L. sativa* cv. Salinas, and shows 75 % identity (82.2 % in the RKIP domain), at the amino acid level to *At*FT, see figure 3.5. To obtain the 5'UTR and promoter sequence along with information regarding exon/intron boundaries this EST was used to probe the lettuce BAC library to obtain the genomic sequence of the lettuce *FT* gene.

One positive hit was identified on screening the library as described in section 3.2.3, and the clone (141M22) was obtained from AGI. The identity of BAC clone 141M22 was verified by PCR using primers designed to the EST

CLSS2320.b1_O03.ab1 sequence; FT_EST_F1 and FT_EST_R1 which produced a product of 332 bp.

BAC DNA was digested using three restriction enzymes, known not to cut within the sequence of EST CLSS2320.b1_O03.ab1 and was Southern blotted. The Southern blot was probed with radioactively labeled EST CLSS2320.b1_O03.ab1. Figure 3.6 shows the resulting blot image.

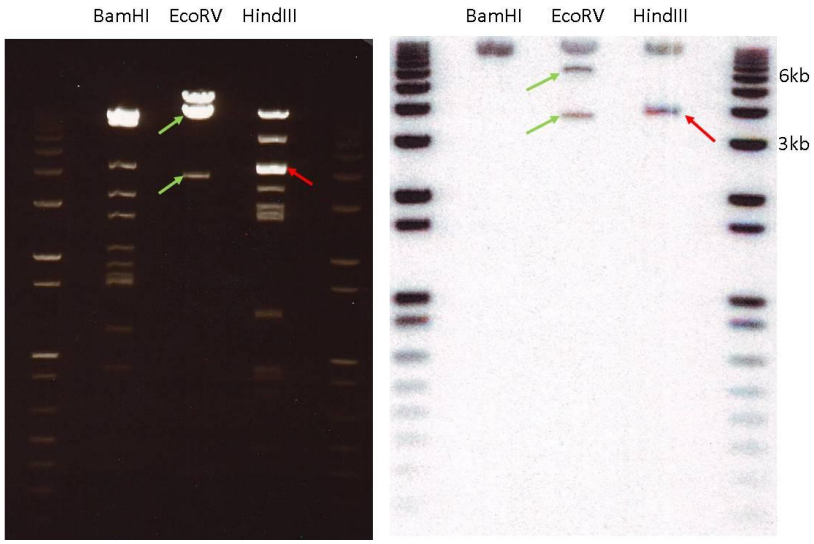


Figure 3.6 - Southern blot results; BAC clone 141M22, digested with *HindIII*, produces a 4.5 kb fragment (red arrow) when probed with EST CLSS2320.b1_O03.ab1. Two fragments are produced when the BAC clone is digested with *EcoRV* (green arrows). Image on left is the fragments generated on a 0.8 % agarose/TAE gel, image on right is the Southern blot of the gel.

Hybridising fragments of ~7.5 and ~4 kb were obtained from the *EcoRV* digest, and a fragment of ~4.5 kb was obtained with *HindIII*. The full length *AtFT* gene is 2483 bp in length, so the *HindIII* fragment could have potentially contained the full length *LsFT* gene homologue. To confirm that the fragment contained the full EST CLSS2320.b1_O03.ab1 sequence, PCR using two sets of primers designed to either *AtFT* exon 1-2 and exon 4 was carried out on the isolated *HindIII* DNA fragment. FT_EST_F1 and R1 produced an expected product of 332 bp and FT_exon4_F and FT_exon4_R produced an expected product of 199 bp.

As PCR confirmed the presence of what appeared to be the full length *LsFT* coding sequence, sub-cloning of this *HindIII* fragment was attempted. However our collaborators in Professor Richard Michelmore's lab at UC Davis, also identified the same clone of interest when screening the BAC library with *AtFT*, and they sequenced the whole of BAC clone 141M22. This sequence was made available to us and so the full length *LsFT* gene sequence was obtained.

LsFT, like *AtFT*, is comprised of four exons and a coding sequence of 527 bp, encoding 176 amino acids. All four exons are similar in size, however the intron sizes do differ, see figure 3.7.

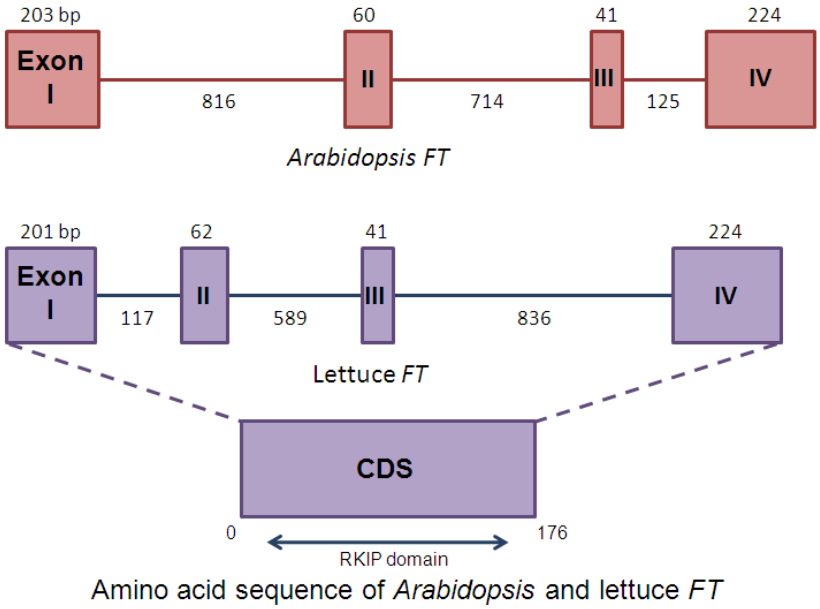


Figure 3.7 – Structure of *Arabidopsis* and lettuce *FT* genomic and coding sequence. *FT* in both *Arabidopsis* and lettuce is encoded by 176 amino acids. Exons are shown as boxes, introns are shown as line. (This applies to all figures).

Upon BLASTing the *Ls*FT protein coding sequence, the top hits were all related to *FT* genes in many different plant species. Figure 3.8 shows a phylogenetic tree constructed to illustrate how *LsFT* is highly conserved to *FT* genes / putative *FT* genes identified in other plant species. The tree is constructed based on the highly conserved PEBP domain. The tree is made up of three distinct groups, this agrees with previous data also based on the PEBP domain published by Faure *et al*., (2007).

LsFT is found within Group I, which includes *AtFT*, rice *FT* orthologue, *Hd3a*, grapevine, *VvFT* (Carmona *et al*., 2007) and a putative *FT* orthologue in *Antirrhinum* (Dr. Massiah, personal communication). It also illustrates the significant similarity between *FT* genes and *AtTSF*, as well as the differences between *FT* genes and *AtMFT*.

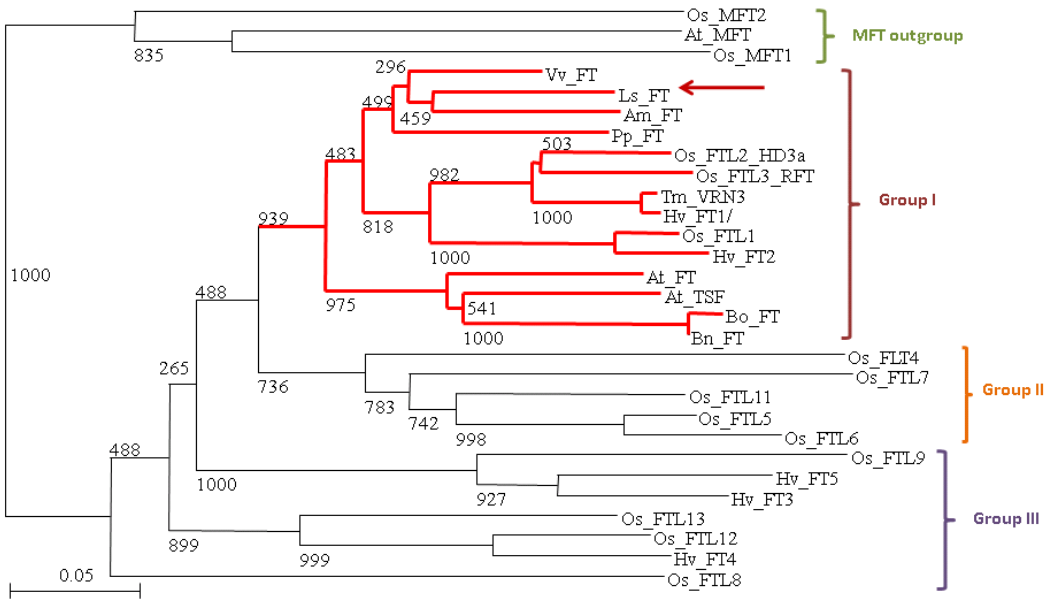


Figure 3.8 – Phylogenetic analysis of *AtFT* compared with *AtTSF*, *AtMFT* and *FT*-like sequences from other plant species. The red arrow highlights *LsFT* which sits in a clade with *AtFT* and putative *FT* genes from other species including grape (*Vv*) and snapdragon (*Am*). The numbers represent bootstrap values from 1000 replicates. Am – *Antirrhinum majus*, At – *Arabidopsis thaliana*, Bn – *Brassica napus*, Bo – *Brassica oleracea*, *Hv* – *Hordeum vulgare*, Ls – *Lactuca sativa*, Os – *Oryza sativa*, Pp – *Prunus persica* Tm – *Triticum monoccum*, Vv - *Vitis vinifera*. (Accession number details can be found in Appendix V, table A4).

3.3.2.3 Investigating the function of *LsFT*

To demonstrate the function of *LsFT*, complementation experiments in *Arabidopsis* were performed, as described in section 3.2.6. The late-flowering *Arabidopsis* mutant, *ft-1* was transformed with a construct over-expressing the *LsFT* coding sequence under the control of the *CaMV 35S* promoter. The *LsFT* gene was cloned into the pB2GW7 vector, which contains a gene that confers resistance to BASTA. The T1 seed from florally dipped plants were selected with BASTA, only plants transformed with the construct conferring resistance grew. *ft-1* is in the Landsberg

erecta, (L*er*) background (Koornneef *et al*., 1991), so *Ler* WT plants were also transformed with the *LsFT* gene, to investigate any change in flowering time. WT L*er* and *ft-1* plants were also transformed with *AtFT* as a positive control.

Plants were scored under a LD photoperiod, non-transformed WT L*er* plants flowered with an average of 7.9 +/-0.32 leaves, compared to non-transformed *ft-1* plants which flowered with an average of 22 +/-0.48 leaves. Very little variation in the number of leaves at flowering was observed in each line of T1 transformants. Mutant plants transformed with *AtFT* and *LsFT* flowered significantly earlier than *ft-1*, and earlier or around the same time as L*er* WT plants, with an average of 4 +/-0.18 (p=<0.001; d.f.=22; l.s.d.=1.469) and 8.3 +/-0.6 (p=<0.001; d.f.=19; l.s.d.=1.341) leaves respectively. Furthermore, ANOVA showed that L*er* plants transformed with *AtFT* (p=<0.001; d.f.=19; l.s.d.=0.621) or *LsFT* (p=<0.001; d.f.=23; l.s.d.=0.754) both flowered significantly earlier, than L*er* WT plants, with an average of 4.4 +/-0.29 and 6.2 +/-0.31 leaves respectively, see figure 3.9.

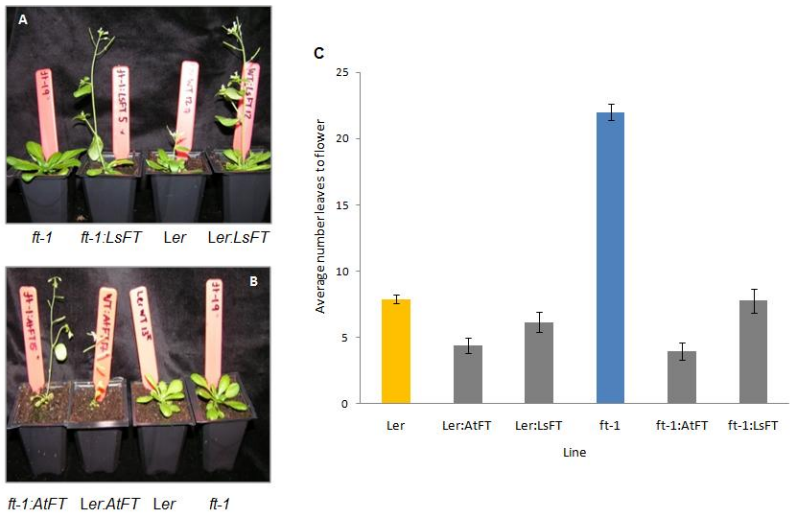


Figure 3.9 – *FT* complementation – T1 transformants
(A) *LsFT* – Examples of *ft-1*, *ft-1:LsFT*, L*er*, L*er:LsFT* plants (B) *AtFT* – Examples of *ft-1:AtFT*, L*er:AtFT*, L*er*, *ft-1* plants (C) Average number of leaves to flower for each of the transformed lines compared to L*er* WT and *ft-1*

The BASTA resistant primary transformant (T1) plants were screened at the molecular level for the presence of the *LsFT* transgene. All transformed plants contained the 269 bp product amplified using *LsFT* specific primers FT_EST1_F1 and FT_exon2-3_R. No product was seen in non-transformed L*er* and *ft-1* control plants. T2 seed was collected from the T1 plants. T2 seed was sown from five individual T1 *ft-1:LsFT* plants. The plants were scored for flowering under both LD and SD photoperiods. Each of the five *ft-1:LsFT* T2 lines flowered significantly earlier than the mutant *ft-1* plants screened under both LD (p=<0.001; d.f.=254; l.s.d.=0.549) and SD (p=<0.001; d.f.=105; l.s.d.=3.31) conditions.

If there is a single transgene segregating in the T2 population a 3:1 ratio of plants flowering early to plants flowering as *ft-1* would be expected from the *ft-1:LsFT* population. Table 3.2 summarises the data obtained from this experiment. Chi-squared analysis of the segregation ratios observed in these experiments was carried out. The progeny from three of the five T1 *ft-1:LsFT* plants (6, 8 and 10), grown under a LD photoperiod flowered in a significant 3:1 ratio, (x^2=0.006, 0.44 and 0.005 respectively). Lines 3 and 4 have a skewed ratio and were shown by to not be segregating in a 3:1 ratio (x^2=6.84 and 7.02 respectively), suggesting that there may be multiple T-DNA insertions in these lines. However all five of the T2 lines grown under SD photoperiods flowered in a significant 3:1 ratio, see table 3.2 for details.

T2 seed collected from five L*er:LsFT* T1 plants was also screened, the plants flowered significantly earlier (7.4-8.1 leaves) than L*er* WT (9.8 +/-0.15 leaves) under both LD (p=<0.001; d.f.=386; l.s.d.=0.426) and SD (p=<0.001; d.f.=47; l.s.d.=1.43) conditions, with L*er* WT plants flowering with 20.8 +/-1 leaves and the T2 plants flowering with 12.5-18.8 leaves.

LONG DAYS						
T2 Line	**Number of T2 plants**	**Plants flowering earlier than *ft-1***		**Plants flowering as *ft-1***		**x^2**
		(%)	**Ave. fl'ing time (leaves)**	**(%)**	**Ave. fl'ing time (leaves)**	
***ft-1*:*LsFT*8**	48	70.83	8.2	29.14	16.5	0.44*
***ft-1*:*LsFT*10**	73	75.34	8.9	24.66	16.5	0.005*
***ft-1*:*LsFT*4**	86	87.21	9	12.79	16	6.84
***ft-1*:*LsFT*3**	65	89.23	8.3	10.77	16.3	7.02
***ft-1*:*LsFT*6**	53	75.47	8.7	24.53	16	0.006*
L*er*	41	100	9.7	0	N/A	N/A
ft-1	20	0	N/A	100	15.8	N/A

SHORT DAYS						
T2 Line	**Number of T2 plants**	**Plants flowering earlier than *ft-1***		**Plants flowering as *ft-1***		**x^2**
		(%)	**Ave. fl'ing time (leaves)**	**(%)**	**Ave. fl'ing time (leaves)**	
***ft-1*:*LsFT*8**	25	76	13.8	24	34.8	0.02*
***ft-1*:*LsFT*10**	30	75.86	13.3	24.14	37.8	0.01*
***ft-1*:*LsFT*4**	17	70.59	14.2	29.41	35.2	0.18*
***ft-1*:*LsFT*3**	32	84.38	14.7	15.62	37.7	1.5*
***ft-1*:*LsFT*6**	22	77.27	14.4	22.73	38.2	0.06*
L*er*	14	100	19	0	N/A	N/A
ft-1	16	0	N/A	100	37.3	N/A

Table 3.2 – Segregation of the flowering phenotype in T2 *Arabidopsis* mutant *ft-1* plants that have been transformed with *LsFT*. Seed generated from five T1 plants was screened for number of leaves at flowering. x^2 = Chi-squared analysis of the 3:1 expected segregation ratio; * = significant 3:1 flowering ratio at 5% significance level.

Genomic DNA was extracted, as described in section 2.2.2, from 15 T2 plants showing a range of flowering times from both *ft-1*:*LsFT* lines 8 and 10. *LsFT* specific primers FT_EST1_F1 and FT_exon2-3_R were used to test for the presence of the transgene, a 269 bp product was amplified from plants which flowered with 6-10 leaves; the number of leaves produced when L*er* plants flowered, see figure 3.10. However plants which flowered with 14-17 leaves, similar to the number observed when *ft-1* plants flowered, did not produce the expected PCR product, indicating that they did not contain the transgene. This shows that the transgene is segregating in the T2 population and is associated with the early flowering phenotype, see figure 3.10. The lane of the gel photo labeled G corresponds to the 976 bp PCR product

amplified from Larissa WT genomic DNA, this product includes sequence from exon1 to 3, including intronic sequence, and is therefore larger in size.

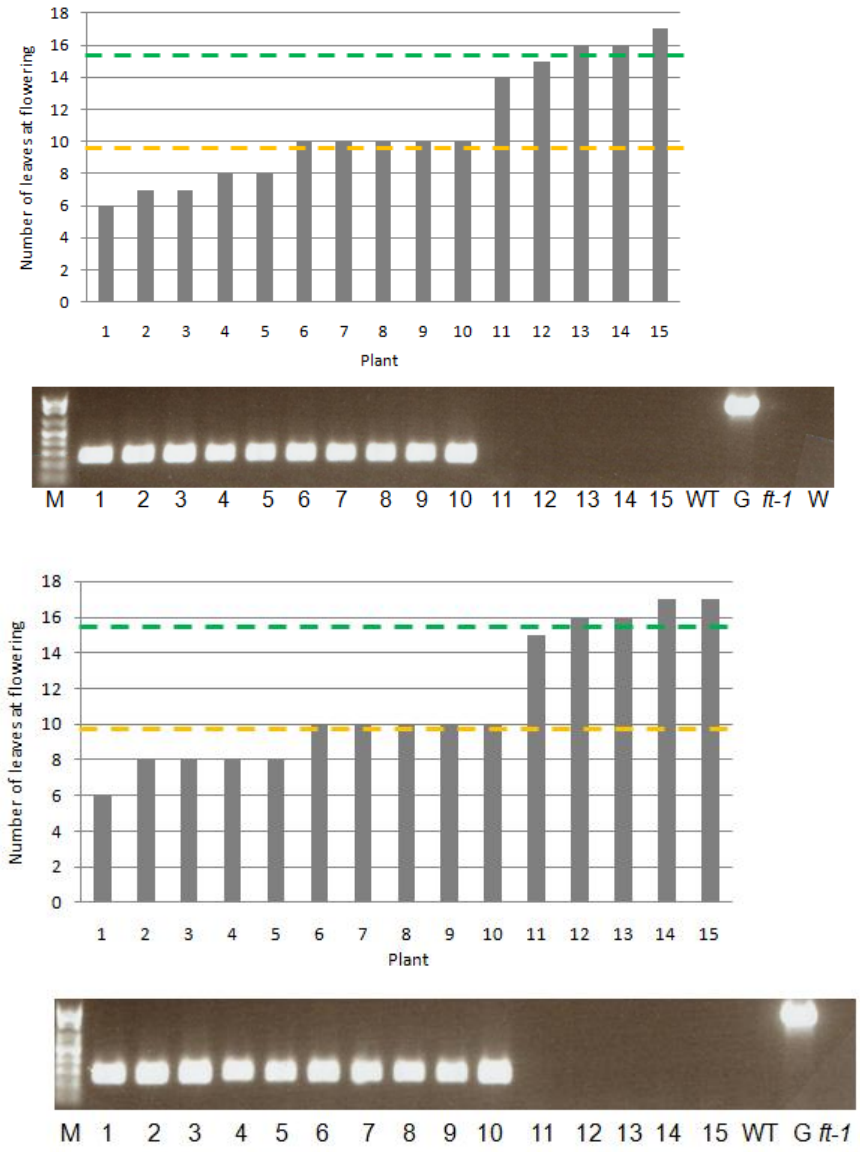


Figure 3.10 – *Arabidopsis FT* T2 transformations scored under LDs. Data obtained from 15 seeds sown from two independent primary transformants, *ft-1:LsFT*8 and *ft-1:LsFT*10. Plants flowering as L*er*, with 6-10 leaves, contain the transgene, (a 269 bp PCR product), however plants flowering with a similar number of leaves to the *ft-1* mutant, 14-17 leaves do not contain the transgene. The average number of leaves at flowering in L*er* plants is highlighted with an orange line, while the average number of leaves at flowering in the *ft-1* mutant plant is highlighted with a green line. M = 1 kb+ marker, G = Larissa WT genomic DNA and WT = L*er*.

Therefore it can be concluded that over expression of the *LsFT* gene is able to rescue the late flowering phenotype of the *ft-1* mutant, suggesting that *LsFT* can perform the same function as *AtFT* and is thus a lettuce orthologue of *AtFT*.

3.3.3 FLK

3.3.3.1 Introduction

Arabidopsis FLK (*AtFLK*), was identified as an autonomous pathway gene through the analysis of the *flk-1, flk-2, flk-3* and *flk-4* mutants (Lim *et al*., 2004; Mockler *et al*., 2004). These mutants result in a late flowering phenotype that can be reversed with a vernalisation treatment. An increase in *FLC* transcripts, and a corresponding decreased level of *FT* and *SOC1* transcripts are observed in the mutant plants which explains the late flowering phenotype.

*At*FLK is a plant specific, predominantly nuclear protein that contains three K-homology (KH)-type RNA-binding domains (Lim *et a*l., 2004; Mockler *et al*., 2004), related to the poly(rC)-binding protein (PCBP) group of heterogeneous nuclear ribonucleoproteins (Makeyev and Liebhaber, 2002). *AtFLK* is made up of six exons, encoded by 577 amino acid residues. The KH domains are located as follows; KH domain I covers residues 188-251, KH domain II – 278-346 and KH domain III – 459-522, see figure 3.11. The KH domains provide the structural basis for RNA-protein and protein-protein interactions (Grishin, 2001; Valverde *et al*., 2008); mutations within these domains have been shown to cause dramatic phenotypic alterations in animals (Liu *et al*., 2008). The gene was designated *FLK* by Lim *et al*., 2004, standing for Flowering Late with KH motifs.

3.3.3.2 Isolation of lettuce *FLK* (*LsFLK*)

EST QGF19F22.yg.ab1, an EST obtained from *L. serriola* identified in CGP1, displayed the highest level of homology to *At*FLK; 60 % identity was observed over a region of 248 amino acids. The EST was obtained and further sequence was obtained using M13_F and M13_R primers. The EST sequence encoded 328 amino

acid residues covering *At*FLK amino acids 204-531, the sequence displayed 57.8 % identity to *At*FLK. Figure 3.11 shows *At*FLK compared with EST QGF19F22.yg.ab1, the sequence covers the end of the first KH domain and all of KH domains II and III.

Figure 3.11 – Alignment of *At*FLK with the lettuce EST QG19F22.yg.ab1. The green arrows correspond to the three KH domains which have been identified in the characterisation of *AtFLK*

The *Arabidopsis* KH gene family comprises 26 members. EST QGF19F22.yg.ab1 displays a higher level of homology to *At*FLK (57.8 % identity), than it does to the next most closely related *Arabidopsis* gene, the *AtFLK* paralogue *PEPPER* (*PEP*), (44.3 % identity). To obtain the full length gene sequence EST QGF19F22.yg.ab1 was used to probe the lettuce BAC library. Three hits were obtained corresponding to BAC clones 079D19, 106H23 and 136B05. The clones were obtained from AGI and were verified by PCR to ensure they contained sequence represented by EST QGF19F22.yg.ab1 which was used to probe the BAC library. A PCR product of 194 bp was obtained from BAC clones 079D19 and 106H23 using primers FLK_EST1_F and FLK_EST1_R designed based on the EST QGF19F22.yg.ab1 sequence. It was not surprising that no PCR products were amplified from BAC clone 136B05, the hit

representing this clone on the BAC library filter, was very weak and did not exactly match the spotting pattern expected, see figure 3.2 for details of spotting patterns.

BAC DNA was extracted from clones 079D19 and 106H23, the DNA was digested and Southern blotted. Both clones produced similar banding patterns when probed with EST QGF19F22.yg.ab1, see figure 3.12, suggesting that the two clones contain similar regions of the lettuce genome.

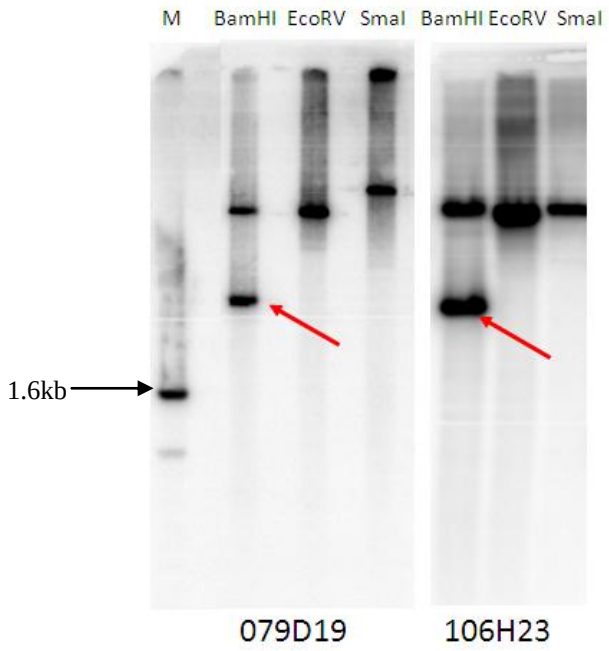


Figure 3.12 – Southern blots of digested BAC clones 079D19 and 106H23, probed with EST QGF19F22 which also shows homology to *AtFLK*. BAC DNA digested with *BamHI* produces a 3.5 kb fragment (red arrow). M = 1 kb+ marker.

The 3.5 kb *BamHI* fragment from BAC 079D19 was sub-cloned as described in section 3.2.5. Sequencing of this fragment showed that it covered most of the EST QGF19F22.yg.ab1, used to probe the BAC library. The sequence covered a region spanning the second exon to the fourth intron of *AtFLK*, however a *BamHI* site was present 48 bp into the fourth intron of the *LsFLK* sequence, meaning no further 3' sequence could be obtained from this fragment. The remainder of the 3'end of the *LsFLK* was obtained using the Genome Walking method as described in section 2.2.7. A ~1.8 kb *DraI* genome walk product was cloned as described in section

2.2.8, the plasmid DNA was sequenced using FLK_GSP2, FLK_GW_seq_exon5_F and FLK_intron5_F primers. Further 5' sequence present in the 3.5 kb *BamHI* fragment was obtained by sequencing plasmid DNA using primers FLK_R_intron2seq and FLK_R_intron2_2seq, approximately 2 kb of extra sequence, covering 1052 bp of 5' UTR was obtained.

The full length *LsFLK* sequence covers 3973 bp. *Ls*FLK is made up of 484 amino acid residues encoded by six exons, interestingly *At*FLK is also encoded by six exons. A direct comparison of the structure of both genes is made in figure 3.13.

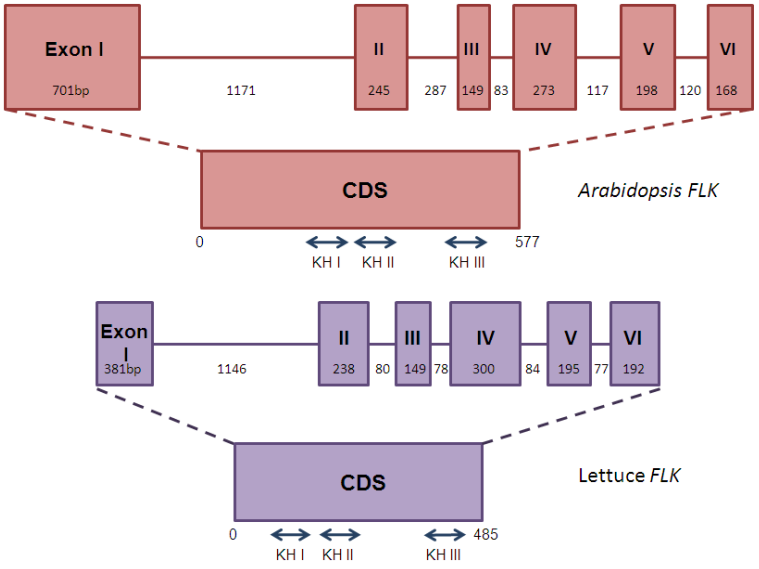


Figure 3.13 – *Arabidopsis* and the putative lettuce *FLK* gene structure. *FLK* is encoded by six exons in both species; *Arabidopsis FLK* is 577 amino acids compared with 485 amino acids in lettuce.

*At*FLK is considerably larger in size at 577 amino acids with the main difference between *LsFLK* and *AtFLK* in the size of the first exon. However, sequences of homologues of *At*FLK in other plant species, including rice and grape, show that they also have smaller first exons than AtFLK, with the remainder of the coding sequence being highly conserved in size. The N-terminal region of *AtFLK* shows no sequence homology to known eukaryotic proteins (Lim *et al*., 2004). Upon BLASTing *Ls*FLK the top *Arabidopsis* hit is FLK. The overall level of identity between the two genes is 55.6 %, this is significantly higher than 40.3 % identity to *At*PEP, *LsFLK* also

contains three KH domains, and is 72.4 % similar to *AtFLK* in these regions, see figure 3.14.

Figure 3.14 – Amino acid sequence comparison of the three KH-domains, highlighted with green arrows, of *At*FLK and *Ls*FLK. *Ls*FLK displays 72.4 % identity to *At*FLK when comparing the three KH domains.

3.3.3.3 Investigating the function of *LsFLK*

As with *LsFT* the function of *LsFLK* was analysed using complementation experiments in *Arabidopsis*. The late-flowering *Arabidopsis* mutant, *flk-4* was transformed with a construct over-expressing the *LsFLK* coding sequence under the control of the 35S promoter as explained for *LsFT* in section 3.2.6 and 3.3.2.3. *flk-4* is in the Columbia, (Col-0) background, so Col-0 WT plants were also transformed with the *LsFLK* gene, to investigate any change in flowering time.

Plants were scored under a LD photoperiod, non-transformed WT Col-0 plants flowered with an average of 9.5 +/-0.3 leaves, compared to non-transformed *flk-4* plants which flowered with an average of 16.1 +/-1.2 leaves. Mutant plants transformed with *LsFLK* flowered significantly earlier than *flk-4*, with an average of 11 +/-0.56 leaves (p=<0.001; d.f.=10; l.s.d.=0.784). Furthermore, ANOVA analysis showed that Col-0 plants transformed with *LsFLK*, which flowered after an average of 6.5 +/-0.24 leaves, flowered significantly earlier (p=<0.001; d.f.=7; l.s.d.=1.018), than Col-0 WT plants, see figure 3.15. The presence of the transgene in the

complemented lines was confirmed by PCR; a 151 bp product was amplified from genomic DNA extracted from each of the T1 transformants using FLK_exon2_F and FLK_EST1_R.

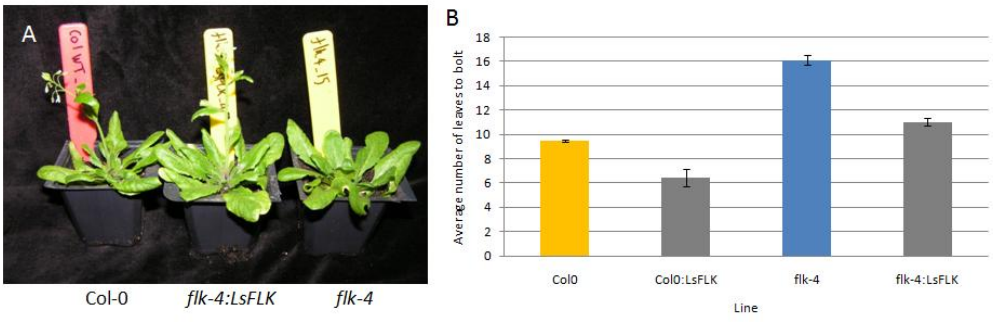


Figure 3.15 – *FLK* complementation – T1 transformants
(A) A *flk-4*:*LsFLK* transformant with a restored flowering phenotype compared with Col-0 and *flk-4*
(B) Average number of leaves to flower for the transformed lines compared to Col-0 and *flk-4*

T2 seed was collected from the T1 plants screened, T2 plants were grown under a LD photoperiod and were scored for the number of leaves at flowering. Table 3.3 shows a breakdown of the flowering data obtained from T2 Col-0:*LsFLK* plants that were collected from four T1 plants. Each of the four Col-0*:LsFLK* flowers significantly earlier than the Col-0 WT plants screened (p=<0.001; d.f.=146; l.s.d.=0.4415). If there is a single transgene segregating in the T2 population a 3:1 ratio of plants flowering early to plants flowering as Col-0 would be expected from the Col-0*:LsFLK* plants. Chi-squared analysis showed that the progeny from all four of the lines (1, 4, 5 and 7) screened flowered in a significant 3:1 ratio (x^2 = 0.20, 0.11, 0.09 and 1.09 respectively).

LONG DAYS						
T2 Line	Number of T2 plants	Plants fl'ing earlier than Col-0		Plants fl'ing as Col-0		x^2
		(%)	Ave. flowering time (leaves)	(%)	Ave. flowering time (leaves)	
Col-0:*LsFLK1*	41	78.05	6.66	21.95	9.11	0.20*
Col-0:*LsFLK4*	48	77.08	7	22.92	9.64	0.11*
Col-0:*LsFLK5*	56	73.22	7.41	26.78	9.43	0.09*
Col-0:*LsFLK7*	44	81.82	7.03	18.18	9.38	1.09*
Col-0	17	N/A	9.47	100	9.47	N/A

Table 3.3 – Segregation of the flowering phenotype in T2 Col-0 WT plants that have been transformed with *LsFLK*. Seed generated from four T1 plants was screened for number of leaves at flowering. x^2 = Chi-squared analysis of the 3:1 expected segregation ratio; * = significant 3:1 flowering ratio at 5% significance level.

Genomic DNA was extracted, as described in section 2.2.2 from seven T2 plants showing a range of flowering times from Col-0:*LsFLK*4. PCR primers (FLK_exon2_F and FLK_EST1_R) specific to *LsFLK* were used to test for the presence of the transgene, a product of 151 bp was amplified from the plants which flowered earlier than Col-0 WT plants with 6-8 leaves, see figure 3.16A. Plants 5 and 15 however flowered with a similar number of leaves to Col-0 WT and did not contain the transgene. It appears that the transgene is segregating in the T2 population and is associated with the early flowering phenotype

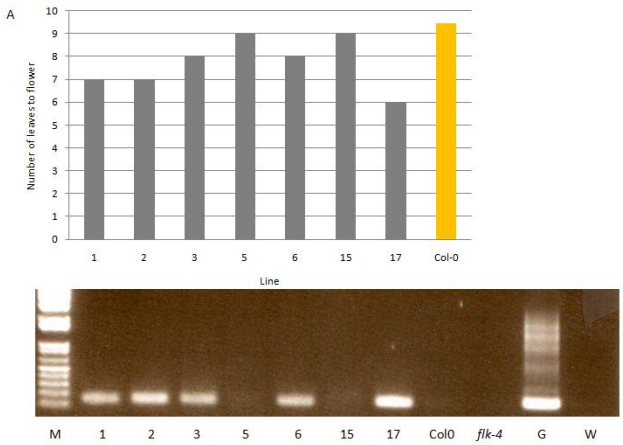


Figure 3.16 – *Arabidopsis FLK* T2 transformations scored under LDs. Data obtained from seven seeds collected from a single T1 primary transformant, Col-0:*LsFLK*4. Plants flowering earlier than Col-0 WT, with 6-8 leaves contain the transgene, (a 151 bp PCR product), however plants flowering with a similar number of leaves to Col-0 with 9 leaves do not contain the transgene. M = 1 kb+ ladder, G = Larissa WT genomic DNA.

T2 seed was also collected from two T1 *flk-4:LsFLK* plants. Chi-squared analysis showed that the progeny from both T1 lines screened (1 and 8) flowered in a significant 3:1 ratio (x^2 = 2.16 and 0.01 respectively), see table 3.4.

This data implies that over expression of the *LsFLK* gene is able to rescue the late flowering phenotype of the *flk-4* mutant, suggesting that *LsFLK* can perform the same function as *AtFLK* and is therefore a lettuce orthologue of *AtFLK*.

LONG DAYS						
T2 Line	**Number of T2 plants**	**Plants flowering as Col0**		**Plants flowering as *flk-4***		**x^2**
		(%)	**Ave. flowering time (leaves)**	**(%)**	**Ave. flowering time (leaves)**	
flk-4:LsFLK1	50	84	9.47	16	20.75	2.16*
flk-4:LsFLK8	22	73.9	10.25	26.1	17.3	0.01*
Col0	17	100	9.2	0	N/A	N/A
flk-4	16	0	N/A	100	18.7	N/A

Table 3.4 – Segregation of the flowering phenotype in T2 *flk-4* plants that have been transformed with *LsFLK*. Seed generated from two T1 plants was screened for number of leaves at flowering. x^2 = Chi-squared analysis of the 3:1 expected segregation ratio; * = significant 3:1 flowering ratio at 5% significance level.

3.3.4 FLD

3.3.4.1 Introduction

Arabidopsis FLD (AtFLD), was identified as an autonomous pathway gene through the analysis of the *Arabidopsis* mutant *fld-1* (Sanda and Amasino 1996), which is in the Col-0 background. *fld-1* flowers much later under inductive LD photoperiods than wild-type Col-0; 49.1 leaves compared to 8.9 leaves. However the flowering time of the mutant can be restored by exposing the plant to a prolonged period of vernalisation.

AtFLD is comprised of two exons that encode a protein of 789 amino acid residues, see figure 3.17.

*At*FLD encodes a plant homologue of the human protein, KIAA0601 (Humphrey *et al*., 2001). These proteins contain a region similar to human polyamine oxidase 1, and a predicted polyamine oxidase (amino acid residues 194-620) in *Arabidopsis* (He *et al*., 2003). KIAA0601 is a component of a complex involved in initiating repression of gene expression by deacetylation of histone residues (Hakimi *et al*., 2003). Mutations in *AtFLD* result in hyperacetylation of histones in *FLC* chromatin resulting in up-regulation of *FLC* expression, and extremely delayed flowering (He *et al*., 2003). *AtFLD* also contains a SWIRM domain (amino acid residues 80-168) found in enzymes involved in chromatin remodeling (He *et al*., 2003). Yang and Chou (1999), also showed that *AtFLD* is involved in formation of flowers in *Arabidopsis*. The combination of a SWIRM and a polyamine oxidase region is also found in the *Arabidopsis* genes *LYSINE-SPECIFIC HISTONE DEMETHYLASE (LDL1)* and *LYSINE-SPECIFIC DEMETHYLASE1 (LDL2).* Another autonomous pathway gene, *AtFCA* described in section 3.3.8, requires *AtFLD* to downregulate *AtFLC* through H3K4 demethylation, this results in a subsequent transcriptional silencing of *AtFLC* (Liu *et al*., 2007; Bäurle and Dean, 2008).

3.3.4.2 Isolation of a lettuce *FLD* (*LsFLD*)

EST clone CLRX5844.b2_G21.ab1, a transcript obtained from *L. serriola*, displayed the highest level of identity with *At*FLD, further sequence to that present in the database was obtained by sequencing the EST clone using M13_F, M13_R, FLD_ESTseq, FLD_ESTseq2 and FLD_GW_GSP2 primers. The EST sequence covered 671 amino acids (corresponding to *At*FLD amino acids 7-676) and displayed 71.9 % identity to *At*FLD (but only 57.2 % and 55.2 % identity to LDL1 and LDL2), it also covered the SWIRM domain and the polyamine oxidase domain of *At*FLD.

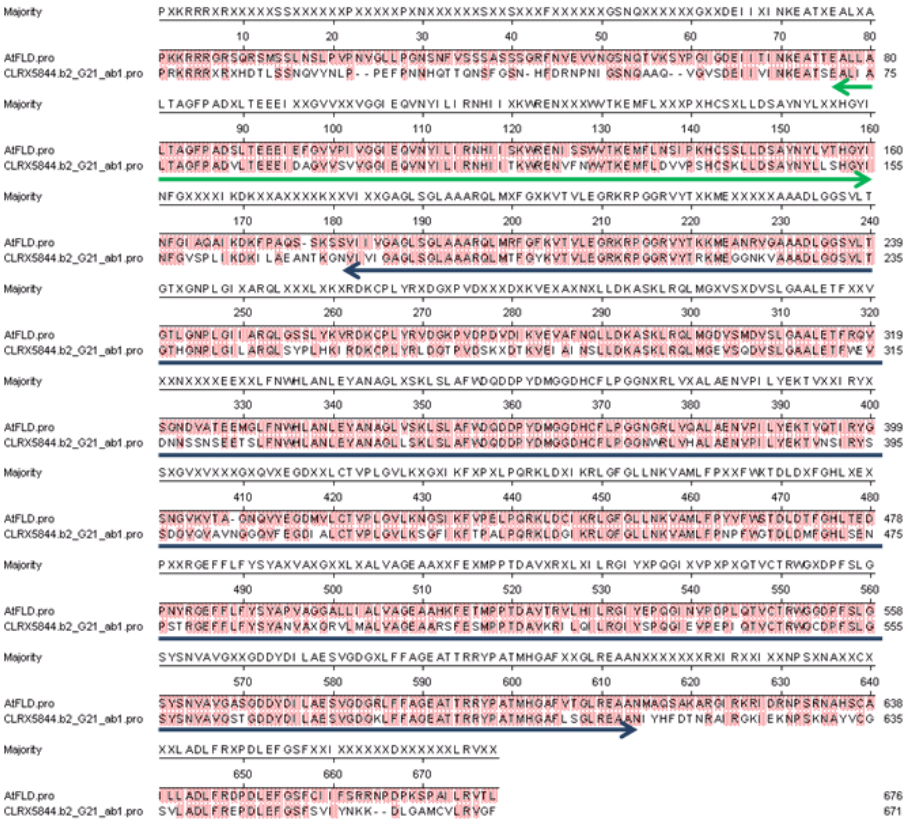


Figure 3.17 – Comparison of *At*FLD coding sequence with EST CLRX5844.b2_G21.ab1; the green arrow indicates the SWIRM domain and the blue arrow the polyamine oxidase domain

Six positive hits were observed from the BAC library screen, figure 3.18 shows a filter indicating three of these hits. Each of the BAC clones was verified, by sequencing PCR products amplified using FLD_EST_1_F and FLD_EST_1_R primers designed to EST CLRX5844.b2_G21.ab1. A PCR product of the expected size of 181 bp was obtained and sequenced from three of the six BAC clones; 151F13, 123I14 and 101C23. Despite showing strong hybridisation, BAC clones 022I17, 097O07 and 007G12 did not produce a PCR product. It is possible that these clones contain other genes with similarity to *FLD*; possibly genes containing SWIRM domains and/or polyamine oxidase domains.

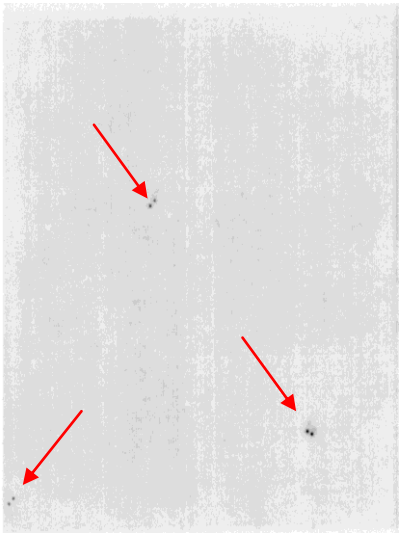


Figure 3.18 – Lettuce BAC library filter probed with EST CLRX5844.b2_G21.ab1 which displays homology to *AtFLD*. The red arrows highlight three hits.

BAC DNA was extracted from each of the three clones which produced PCR products and was digested with *BamHI*, *DraI* and *EcoRV*. The products were Southern blotted and probed using EST CLRX5844.b2_G21.ab1. The probe hybridised to identical sized fragments from each BAC clone. Three *DraI* fragments of interest were identified; ~1.8 kb, ~1.6 kb and ~1.3 kb in size. Sub-cloning of these fragments was attempted, but proved unsuccessful.

Genome walk primers were designed to the 3' and 5' end of the putative *LsFLD* sequence; a fragment of ~1.4 kb was generated using FLD_5_GSP1 and FLD_5_GSP2 primers, from a 5' *DraI* library which covered the remainder of the upstream coding sequence as well as 420 bp of 5'UTR. A further genome walking experiment, utilising the new sequence obtained produced a *PvuII* product of ~1.8 kb, using primers FLD_5_GSP1_2 and FLD_5_GSP2_2. Sequence of ~1.1 kb 5'UTR was verified using FLD_5_GSP2_2 primer. A 3' *PvuII* library product of ~1.6 kb amplified using primers, FLD_3_GSP1 and FLD_3_GSP2, produced sequence covering the remainder of the coding sequence and ~1.1 kb 3'UTR, see figure 3.19 for details of genome walk products obtained.

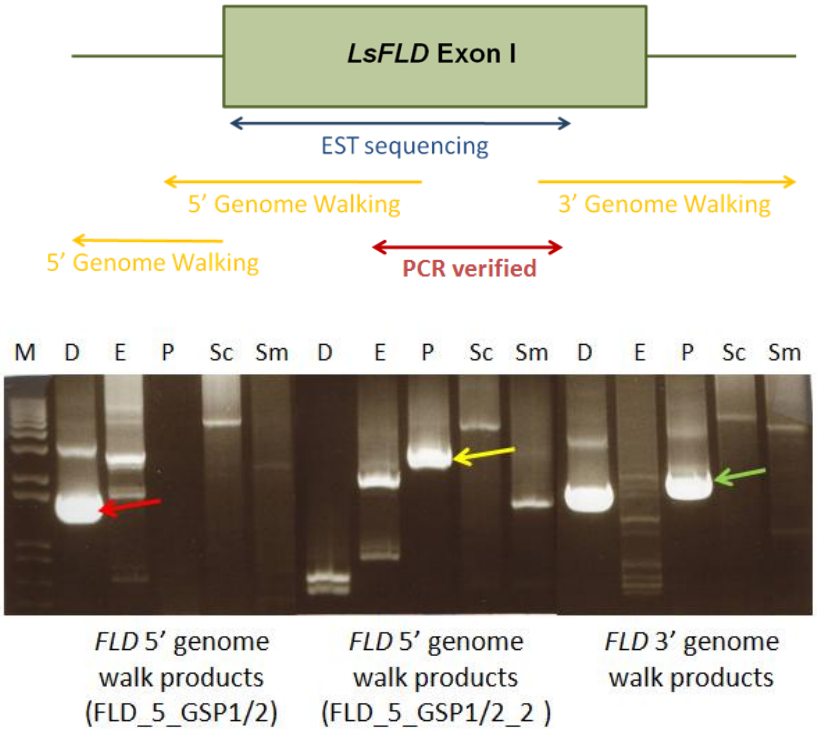


Figure 3.19 – *LsFLD* was obtained by sequencing EST CLRX5844.b2_G21.ab1 and Genome Walking from BAC clone 151F13. The gel image highlights the genome walk products obtained from BAC clone 151F13. The red and yellow arrows highlight *DraI* and *PvuII* 5' gene products respectively, that were cloned and sequenced. The green arrow indicates the *PvuII* 3' gene product that was cloned and sequenced. M = 1 kb+ marker, D = *DraI*, E = *EcoRV*, P = *PvuII*, Sc = *ScaI* and Sm = *SmaI*

Interestingly *Ls*FLD is encoded by one large exon comprising 848 amino acid residues, compared with two exons in *Arabidopsis*, see figure 3.20. To confirm this, PCR primers were designed from the end of the 5' genome walk product to the start of the 3' genome walk product. The product amplified was sequenced, the sequence confirmed that *Ls*FLD contained no introns and was encoded by one exon. Figure 3.19 shows the *LsFLD* gene and highlights the methods used to obtain the sequence.

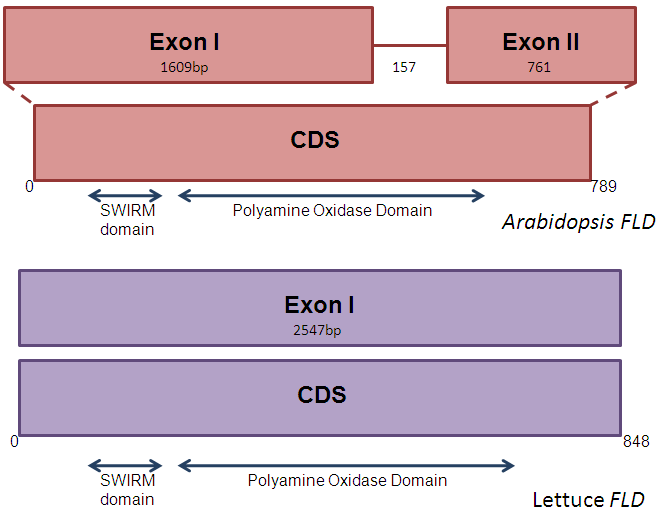


Figure 3.20 – Gene structure of *Arabidopsis FLD* and the putative lettuce *FLD*. *AtFLD* is encoded by two exons compared with a single exon in lettuce; *Arabidopsis FLK* is 789 amino acids compared with 848 amino acids in lettuce.

The top *Arabidopsis* hit when BLASTing the lettuce gene amino acid sequence was FLD, which displays 69.2 % identity. *LsFLD* is also sufficiently different from other *Arabidopsis* genes which contain polyamine oxidase and SWIRM domains, it shows only 52.1 % identity with LDL1and 50.9 % identity LDL2.

3.3.4.3 Investigating the function of *LsFLD*

As with *LsFT* and *LsFLK*, the function of *LsFLD* was analysed using complementation experiments in *Arabidopsis*. The late-flowering *Arabidopsis* mutant, *fld-3* was transformed with a construct over-expressing the *LsFLD* coding sequence under the control of the 35S promoter as explained for *LsFT* is section 3.2.6 and 3.3.2.3. *fld-3* is in the Columbia, (Col-0) background, so Col-0 WT plants were also transformed with the *LsFLD* gene, to investigate any change in flowering time.

Plants were scored under a LD photoperiod, non-transformed WT Col-0 plants flowered with an average of 10.87 +/-0.2 leaves, compared to non-transformed *fld-3* plants which flowered with an average of 39 +/-0.58 leaves. Mutant plants

transformed with *LsFLD* flowered significantly earlier than *fld-3,* with an average of 35.59 +/-0.45 leaves (p=<0.001, d.f.=12, l.s.d.=0.766). However the flowering phenotype was not recovered to a similar number of leaves as recorded for WT, as would be expected should *LsFLD* perform the same function as *AtFLD*. WT flowered with an average of 10.87 leaves, the earliest *fld-3*:*LsFLD* flowered with 31 leaves. This suggests that the transgene has some effect on flowering time when overexpressed in WT plants. However WT flowering is not restored suggesting that *LsFLD* is not the functional orthologue of *AtFLD*, see figure 3.21. A similar result was seen with Col-0 plants that were transformed with the *LsFLD* transgene. The Col-0:*LsFLD* plants flowered with an average of 10.07 +/-0.23 leaves which was similar to Col-0 plants although deemed to be significantly different (p=<0.001, d.f.=29, l.s.d.=0.539). The presence of the transgene within the *Arabidopsis* genome was verified by PCR using genomic DNA extracted from the transformed plants. The expression of the transgene was also confirmed by PCR using cDNA synthesised from the transformed plants (data not shown).

The lack of *LsFLD* complementation observed in *Arabidopsis* may also be due to the difference in plant species. Although the data collected suggests that the *LsFLD* gene is not the functional *AtFLD* orthologue, no experiments have been performed to demonstrate that the *fld-3* mutant can be functionally complemented by the *AtFLD* gene.

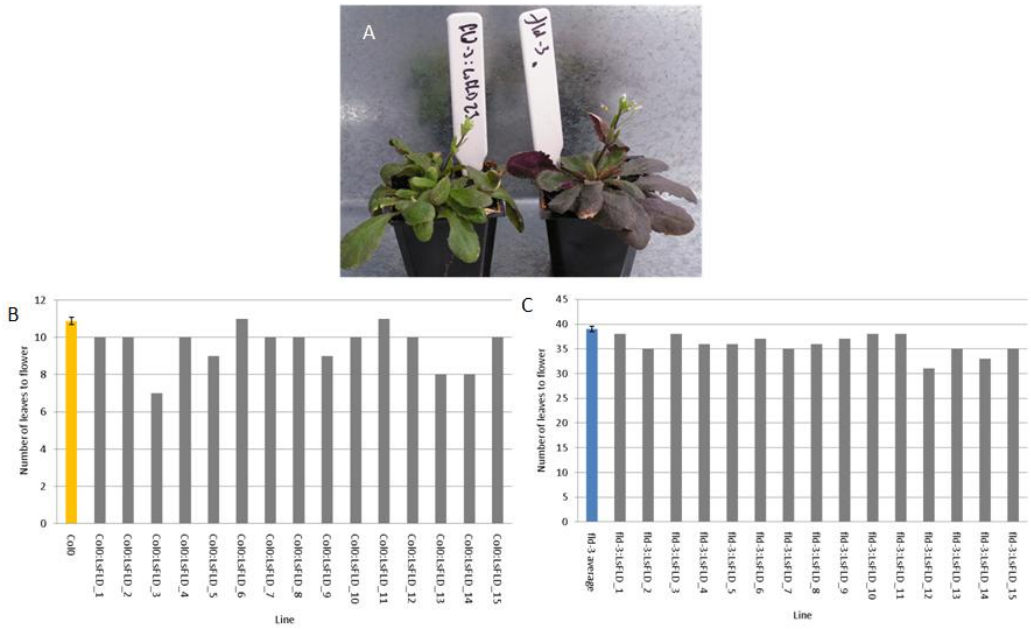


Figure 3.21 – *FLD* complementation – T1 transformants
(A) Example of a *fld-3*:*LsFLD* transformant compared with the mutant *fld-3*
(B) Example of number of leaves to flower for Col-0 WT compared with Col-0:*LsFLD* T1 plants
(C) Example of number of leaves to flower for Col-0 compared with *fld-3:LsFLD* T1 plants

3.3.5 LD

3.3.5.1 Introduction

Flowering in *Arabidopsis LD (AtLD)*, mutants *ld-1* and *ld-3* is delayed when compared to WT under both LD and SD photoperiods, with a vernalisation treatment rescuing the late flowering phenotype (Redei, 1962; Lee *et al*., 1994).

The *At*LD protein encodes a homeodomain associated with DNA binding (Lee *et al*., 1994), and in rare cases it has been shown to be involved in RNA binding (Dubnau and Struhl, 1996). The coding sequence of *AtLD* is 953 amino acid residues in length and as well as the homeodomain (amino acids 66-123) contains two putative bipartite localisation signals and a glutamine-rich region at the carboxy terminus, which resembles glutamine rich domains found in several transcription factors (Mitchell and Tijan, 1989), see figure 3.22. The central region of the *At*LD protein also contains five repeated QPVNG motifs separated by a four-amino acid spacer (Lee *et al*., 1994). *AtLD* is expressed primarily in regions of the shoot and root apex

containing dividing cells, including the apical meristems. The *AtLD* gene product is targeted to the nucleus, this is consistent with its proposed function in regulating transcription (Aukerman *et al*., 1999)

3.3.5.2 Isolation of a lettuce *LD* (*LsLD*)

EST CLSM7821.b1_J11.ab1 was the only EST clone to show any homology to *AtLD*. Sequence comparison of *AtLD* with *Zea mays LD*, (*ZmLD*) by van Nocker *et al*., (2000), identified four regions of homology, see table 3.4, EST CLSM7821.b1_J11.ab1 showed homology to Region I.

	Amino acid residues	
Region	***At*LD**	***Zm*LD**
I	31-130	41-140
II	205-318	280-392
III	395-445	462-512
IV	580-659	600-675

Table 3.5 – Homology identified between *At*LD and *Zm*LD (van Nocker *et al*., 2000), in four regions of the gene coding sequence

Further sequence to that published in the database was obtained from EST CLSM7821.b1_J11.ab1 using M13_R and LD_ESTseq primers. The sequence obtained covered *At*LD amino acid residues 55-478, and is 60.7 % conserved at the amino acid level, this includes the homeodomain region of the protein (*At*LD amino acids 65-133), where the level of identity rises to 66 %, see figure 3.22. The full EST CLSM7821.b1_J11.ab1 sequence also showed high levels of homology to both regions II (76.8 %), and III (62.7 %).

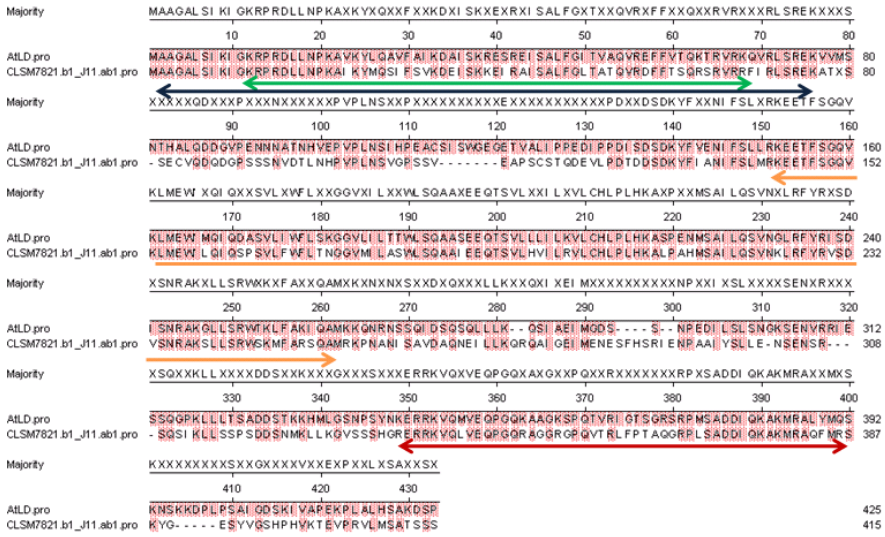


Figure 3.22 – Alignment of *At*LD with EST CLSM7821.b1_J11.ab1. The homeodomain is highlighted with a green arrow. Regions I, II and III of homology between *At*LD and *Zm*LD reported by van Nocker *et al*., 2000, are highlighted with a blue, orange and red arrow respectively

EST CLSM7821.b1_J11.ab1 was used to probe the lettuce BAC library. One hybridisation hit was recorded from the screen, see figure 3.23.

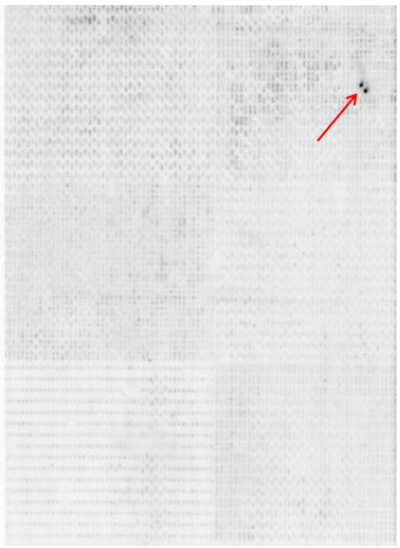


Figure 3.23 - Lettuce BAC library filter probed with EST CLSM7821.b1_J11.ab1 which displays homology to *AtLD*. The red arrow highlights the one hit observed.

This suggests that *LD* may be a single copy gene in lettuce, as has been reported in *Arabidopsis* and Maize (Lee *et al*., 1994; van Nocker *et el*., 2000). The hit obtained was BAC clone 0206I8; DNA was extracted from the BAC clone and was digested

as described in section 3.24 and 3.25. Cloning of two fragments generated from a *HindIII* digest of ~6.5 kb and ~5.5 kb was attempted, but was unsuccessful.

A number of methods were utilised to obtain a full length sequence of the putative *LsLD* gene. Figure 3.24 summarises the methods used.

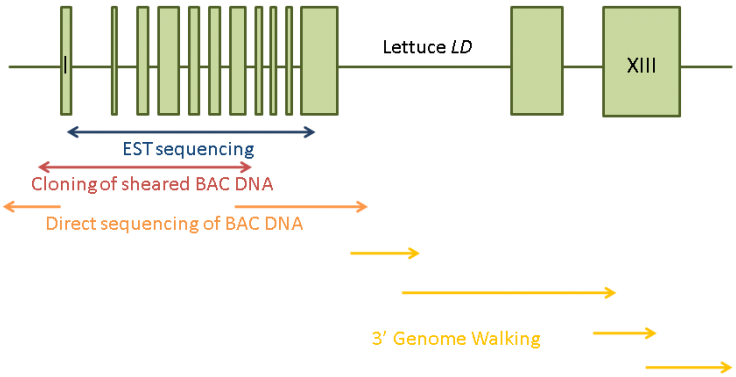


Figure 3.24 – Methods used to isolate a full length lettuce *LD* genomic sequence

Initially a method based on shearing BAC DNA into smaller sized fragments for sub-cloning and sequencing was used to try to obtain full length sequence of *LsLD*, as described in section 3.2.5. This obtained sequence covering *AtLD* exon 1 – intron 7 as well as 318 bp of 5' UTR. In total 2572 bp of sequence was obtained using this method. Direct sequencing of BAC DNA extracted from the clone 0206I8 also provided some sequence information, primers LD_5UTR and LD_5_R_seq2 were used to obtain 800 bp of 5'UTR, meaning 1050 bp of 5'UTR was obtained in total. Direct sequencing of the BAC clone 0206I8 also produced 1248 bp of sequence covering *AtLD* intron7-intron11. The remainder of the *LsLD* gene was obtained by genome walking, four sets of primers were used to walk 4546 bp and this sequence covered *AtLD* from intron11 through to the stop codon at the end of exon 13 plus 600 bp of 3'UTR.

Like *At*LD, the putative lettuce LD gene is also made up of 13 exons, *At*LD is encoded by 953 amino acids compared to 960 encoding *Ls*LD. Figure 3.25,

highlights the differences in the size of the introns and exons encoding the LD gene in each species.

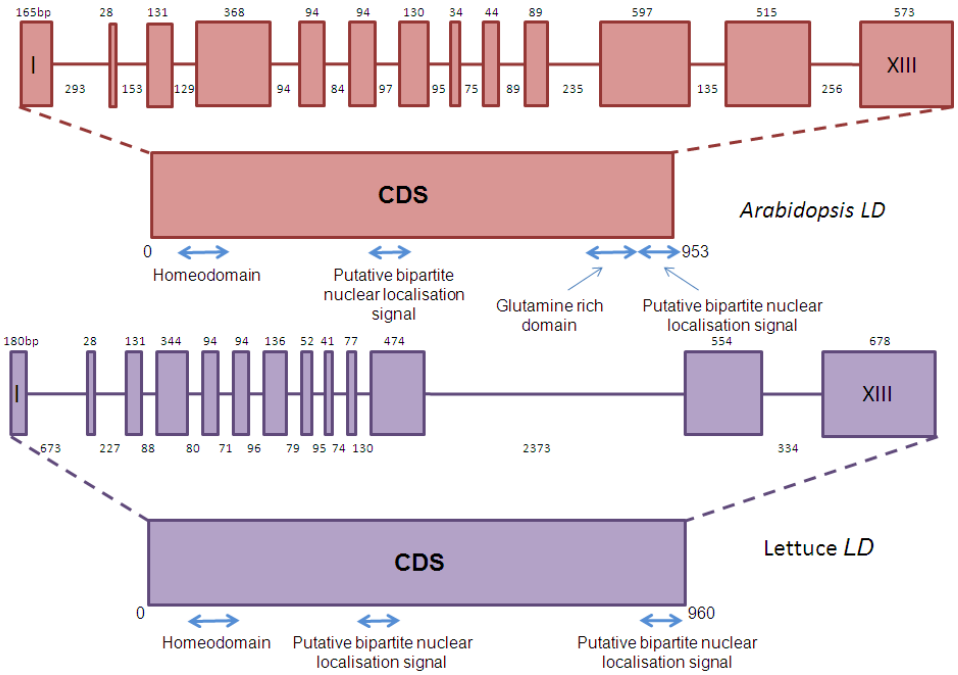


Figure 3.25 – Gene structures of At*LD* and a lettuce *LD* gene homologue, *Ls*LD. Both genes are comprised of 13 exons. The *At*LD contains a glutamine rich region at the C-terminal end of the gene, which is not present in the lettuce homologue.

*Ls*LD displays 47.2 % identity with *At*LD, which is higher than the 42.9 % identity that the putative ZmLD displays with *At*LD. A putative *LD* gene has been identified in grape, this displays a 50.6 % identity with *Ls*LD, see table 3.5 for details.

	***At*LD**	***Ls*LD**	***Zm*LD**	***Vv*LD**
***At*LD**	-	47.2	42.9	45
***Ls*LD**	47.2	-	39.3	50.6
***Zm*LD**	42.9	39.3	-	35.5
***Vv*LD**	45	50.6	35.5	-

Table 3.6 – Comparison of percentage identity between LD in different plant species

Unlike *At*LD and *Zm*LD, *Ls*LD and *Vv*LD do not contain a glutamine rich region at the 3' end of the gene. The two bipartite-type nuclear localisation consensus signals present in the *Arabidopsis* and Maize LD proteins, one of which comprises the final residues of the C-terminal end of the gene (Lee *et al*., 1994; van Nocker *et al*., 2000)

are also found in *Ls*LD. However the five QPVNG repeat sequences observed in *At*LD are not present in *Ls*LD, *Vv*LD or *Zm*LD.

3.3.6 FVE

3.3.6.1 Introduction

Arabidopsis FVE (AtFVE or *AtMSI-4*), is part of the *AtMSI* gene family; five *AtMSI* genes have been characterised with *AtMSI-5* being most closely related to *AtFVE* (78.4 % identity), see figure 3.26.

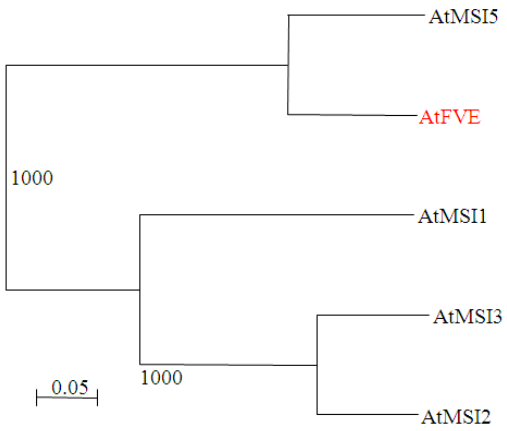


Figure 3.26 – Phylogenetic tree illustrating the *Arabidopsis* MSI gene family. FVE is most similar to MSI-5.

AtFVE is member of the autonomous pathway as demonstrated by Ausin *et al*., (2004). The *AtFVE* mutants, *fve-1, fve-2, fve-3* and *fve-4* all flower later than WT plants in SD and LD photoperiods, the delay is abolished by exposure to 4°C for 12 weeks.

*At*FVE comprises 507 amino acid residues and is made up of 15 exons. The gene contains six WD-40 repeat domains (Kim *et al*., 2004; Ausin *et al*., 2004). The WD-40 repeats cover amino acids 165-415, see figures 3.27 and 3.30.

WD-40 repeats are found in eukaryotic proteins involved in basic cell regulatory processes (Li and Roberts, 2001). *AtFVE* encodes a homologue of the mammalian

retinoblastoma (Rb) associated protein and participates in a protein complex repressing *FLC* transcription through a histone deacetylation mechanism (Ausin *et al*., 2004). Retinoblastoma-associated proteins are present in protein complexes involved in chromatin assembly and histone modification (Ausin *et al*., 2004). In mammals the Rb protein represses transcription by recruiting histone deacetylases involved in repressing gene expression by deacetylation of histone residues (Kim *et al*., 2004). Like *fld* mutants, the *fve* mutant exhibits an increase in *FLC* acetylation; however the increase is not as large as in *fld* mutants (He *et al*., 2003). *AtFVE* is primarily a meristem regulator operating to calibrate growth rates; it is expressed in growing organs where it regulates the timing and speed of differentiation (Morel *et al*., 2009). *fve* mutants have also been shown to have pleiotropic effects on plant architecture and organ production and growth; depending on photoperiod, the vegetative biomass in the mutants is increased three to eight-fold, a larger number of seeds are also produced (Morel *et al*., 2009), this would be a potentially beneficial phenotype should a mutation in *FVE* have the same effect in lettuce.

3.3.6.2 Isolation of a lettuce *FVE* (*LsFVE*)

Numerous ESTs displaying homology to *AtFVE* were found in both CPG1 and CPG2, this is unsurprising as when BLASTing the WD-40 repeat region characterising the *AtFVE* gene 140 hits in *Arabidopsis* alone are returned. However, contig QG_CA_Contig5932 from CPG1, created using five individual ESTs showed 83 % identity at the amino acid level. One of the ESTs making up this contig, QGC24A13.yg.ab1, identified in *L.sativa* cv. Salinas, was sequenced using M13_R and FVE_seqv2, and displayed 78.1 % identity to *At*FVE covering amino acids 48-

260, see figure 3.27. The EST showed limited homology to *At*MSI1-3 (32.3 %-34.1 %), but did display 71.1 % identity to *At*MSI5.

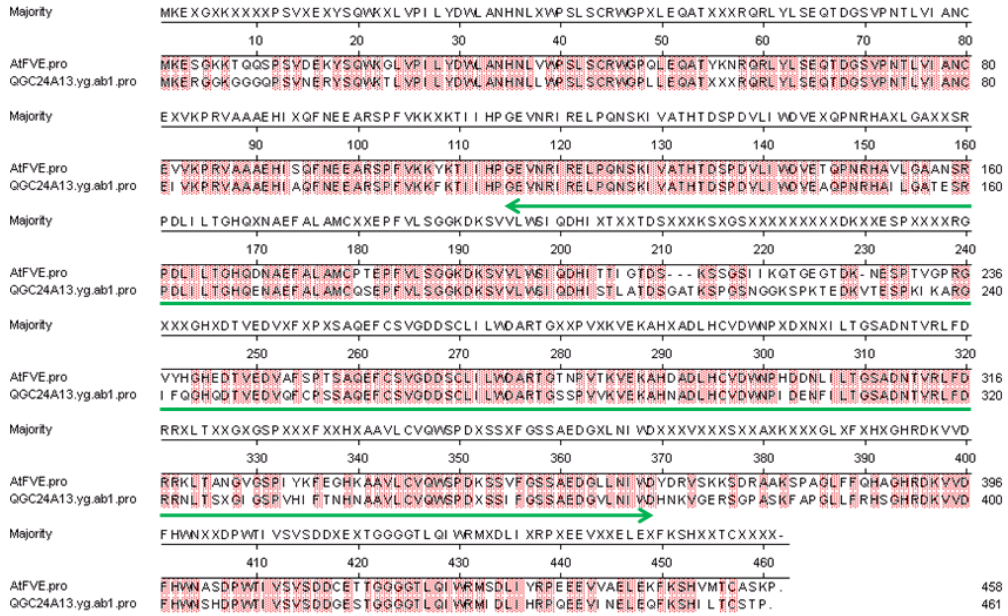


Figure 3.27 – Comparison of EST QGC24A13.yg.ab1 amino acid sequence with *At*FVE sequence. The WD-40 repeat region is underlined in green.

EST QGC24A13.yg.ab1 was used to probe the lettuce BAC library to obtain full length genomic sequence information. Seven hits were recorded from screening the library; they corresponded to BAC clones 162B10, 061A10, 150H09, 058B13, 082D19, 036H17 and 047J02. BAC DNA was extracted from each of the hits. An expected PCR product of 163 bp using primers designed to EST QGC24A13.yg.ab1, (FVE_exon7_F and FVE_exon8_R), was amplified from three of the seven hits. The three BAC clones verified as containing the EST sequence represented the strongest hits from the library screen; the other hits may represent other MSI gene sequences or sequences containing WD-40 repeats. DNA from each of the BAC clones; 036H17, 126B10 and 082D19 was digested and Southern blotted. All three of the clones produced identical banding patterns when probed with EST QGC24A13.yg.ab1. Of particular interest were three fragments of 5 kb, 3.2 kb and 2.3 kb generated following an *EcoRV* digest. Cloning of these fragments was

attempted with no success. A number of methods were used to obtain the full length gene sequence, see figure 3.28.

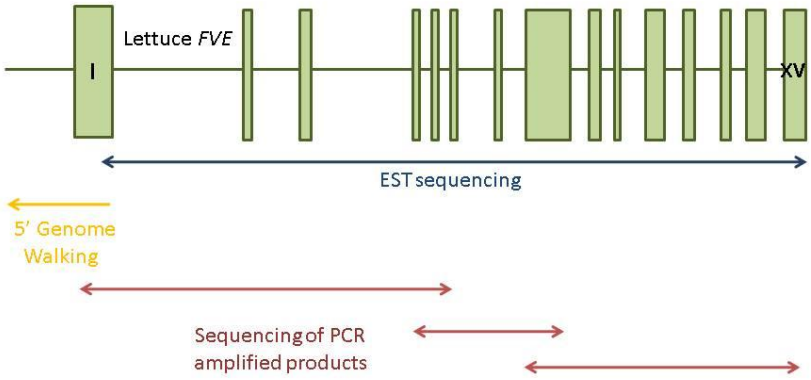


Figure 3.28 - Methods used to isolate a full length lettuce *FVE* genomic sequence

As previously mentioned the EST sequence data did cover most of *At*FVE, from the middle of exon 1 to the stop codon at the end of exon 15. To obtain further exon 1 and 5'UTR sequence of a putative *LsFVE,* genome walking from BAC 036H17 was attempted. Primers FVE_GW_GSP1 and GSP2, were designed to the sequence at the 5' end of EST QGC24A13.yg.ab1. A 1.5 kb *SmaI* GW product was cloned and sequenced, the sequence produced no further homology with *AtFVE*, suggesting the sequence obtained was all 5'UTR and that the sequence obtained from sequencing EST QGC24A13.yg.ab1 represented the full length putative *LsFVE* gene.

PCR primers were designed to obtain the intronic sequence and to confirm intron/exon borders of the *LsFVE* homologue. Products of 1.5 kb, 700 bp and 1.2 kb were amplified using primers covering exon 1 to exon 6, exon 4 to exon 8 and finally exon 8 to exon 15, see figure 3.29, each primer pair overlapped by around 150 bp to ensure full gene coverage. The products were subsequently cloned and sequenced.

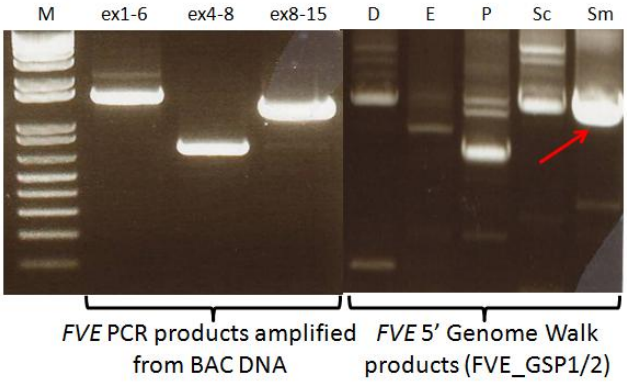


Figure 3.29 – PCR products and genome walk products amplified from BAC clone 036H17, these products were cloned and sequenced. M = 1 kb+ marker, D = *DraI*, E = *EcoRV*, P = *PvuII*, Sc = *ScaI*, Sm = *SmaI* and W = water control.

The full length *Ls*FVE is 460 amino acids in length compared to 507 in *At*FVE, both genes are encoded by 15 exons. As with *Ls*FLK the difference in gene length is due to a shorter first exon, see figure 3.30. Upon BLASTing *Ls*FVE the top hit is *At*FVE, *Ls*FVE displays 79 % identity with *At*FVE, 72.3 % identity with *At*MSI5 and less than 35 % identity with *At*MSI1, 2 and 3.

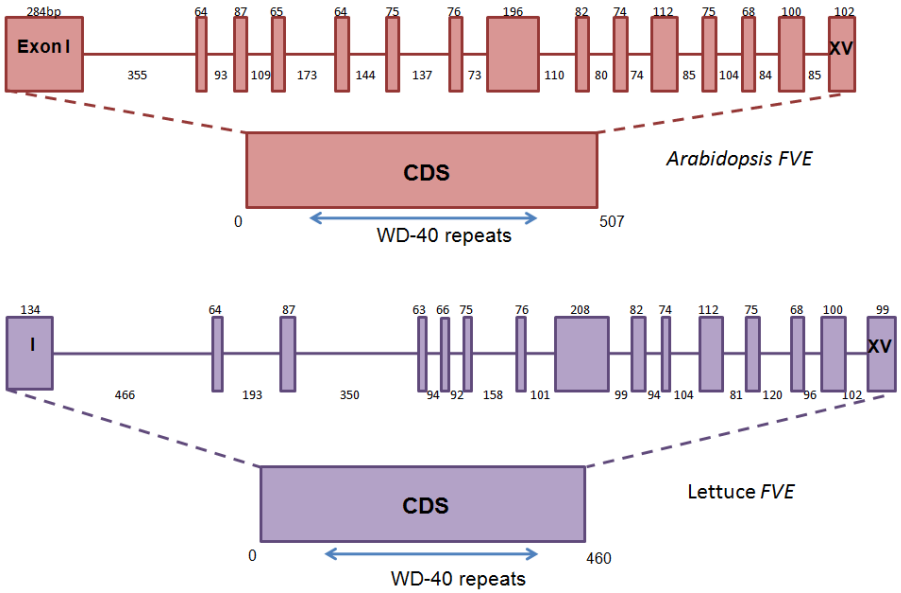


Figure 3.30 - Gene structure of *AtFVE* and a lettuce *FVE* homologue, *LsFVE*. Genes from both species are comprised of 15 similarly sized exons. *At*FVE is slightly bigger at 507 amino acids in length; this is mostly due to a larger exon 1.

Phylogenetic analysis was performed to show how closely conserved it is with *At*FVE and *At*MSI5, see figure 3.31. This tree was constructed based on the WD-40

repeat region of the *Arabidopsis* MSI family genes and *Ls*FVE, *Ls*FVE lies in a group with *At*MSI5 and *At*FVE.

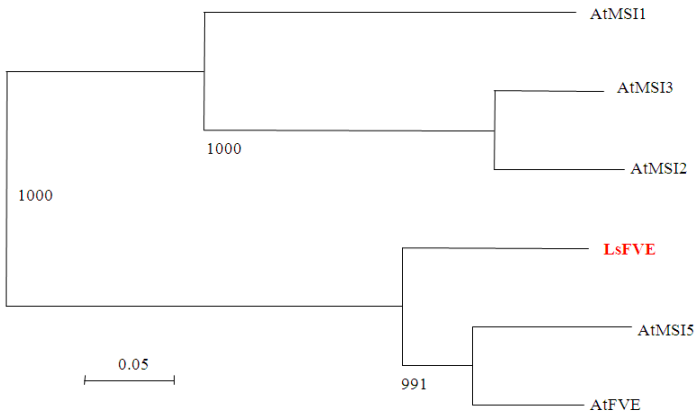


Figure 3.31 – Phylogenetic tree illustrating *Ls*FVE level of homology with *At*FVE, 79 % and *At*MSI5, 72.3 %.

3.3.7 FY

3.3.7.1 Introduction

FY was identified in *Arabidopsis* by searching for proteins which interact with the WW domain of the autonomous pathway gene *FCA*. Simpson *et al*., (2003) showed that FY was required for the negative autoregulation of *FCA* expression. A Pro-Pro-Leu-Pro (PPLP) motif at the C terminal end of *At*FY interacts with the WW domain of *At*FCA, this interaction is essential for *At*FCA function and the down-regulation of *AtFLC* (Simpson *et al*., 2003). This interaction was also shown to be present in rice; *Os*FCA combines with *Os*FY via its WW domain (Jang *et al*., 2009).

*At*FY belongs to a highly conserved group of eukaryotic proteins represented in *S. cerevisiae* by the 3' end-processing factor, Pfs2p. *AtFY* regulates RNA 3'end processing in *Arabidopsis*. *At*FY also contains a conserved region of seven WD-repeats (Simpson *et al*., 2003). The WD-40 repeats cover amino acids 115-405. The flowering phenotypes of *FY* mutants *fy-1*, *fy-2* and *fy-3* have been characterised; plants containing each of the three mutations exhibit a delay in flowering when compared with *Arabidopsis* WT plants. This delay in flowering was restored on each

occasion with a vernalisation treatment, confirming *FY* as a member of the autonomous pathway (Simpson *et al*., 2003; Henderson *et al*., 2005).

3.3.7.2 Isolation of a lettuce *FY* (*LsFY*)

Over 20 transcripts in the lettuce databases exhibited homology to *AtFY.* EST CLPX15054.b1_K19.ab1, a transcript expressed in *L. perennis*, was 203 residues in length and displayed identity at the amino acid level to *At*FY (amino acids 18-199, from the middle of exon 1 – middle exon 5). The EST clone was further sequenced using M13_R and FY_ESTseq_2 primers. The sequence generated extended the database sequence to the stop codon at the end of the 647 amino acids encoding *At*FY, and displayed identity of 70.6 % at the amino acid level. The sequence obtained from EST CLPX15054.b1_K19.ab1 included the WD-40 repeats characterising *At*FY, the level of amino acid identity in this region was 83.2 %, see figure 3.32.

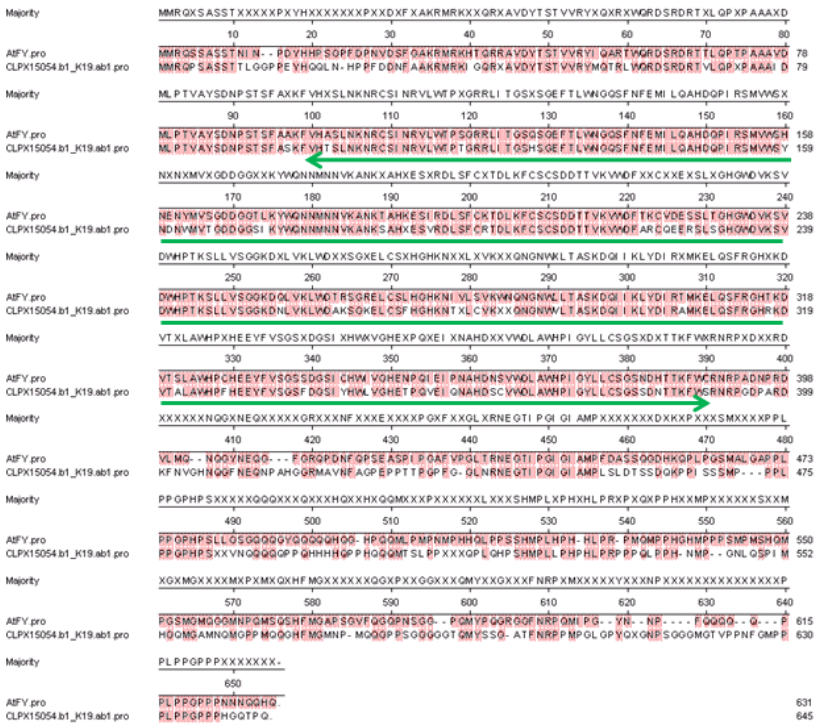


Figure 3.32 – Comparison of lettuce EST CLPX15054.b1_K19.ab1sequence with *At*FY sequence. The WD-40 repeat region is highlighted with a green arrow

EST CLPX15054.b1_K19.ab1 was used to probe the lettuce BAC library, four hits were observed; 008B24, 094D1, 157I13 and 194G5. DNA was extracted from the four clones and PCR amplification of an 81 bp product was carried out using primers FY_EST1_F and FY_EST1_R designed to EST CLPX15054.b1_K19.ab1 to verify the presence of an *LsFY* homologue. Clones 008B24 and 094D1 both amplified the correct sized fragment; no products were amplified from 157I13 and 194G5. Clone 157I13 was a very weak hit, however, 194G5 was a strong hit and may represent a gene closely related to *At*FY, potentially a clone containing WD-40 repeats. Southern blotting produced the same banding pattern in both clones, suggesting the two clones covered the same region of the lettuce genome.

Once again a number of methods, summarised in figure 3.33, were employed to obtain a full length *Ls*FY gene homologue.

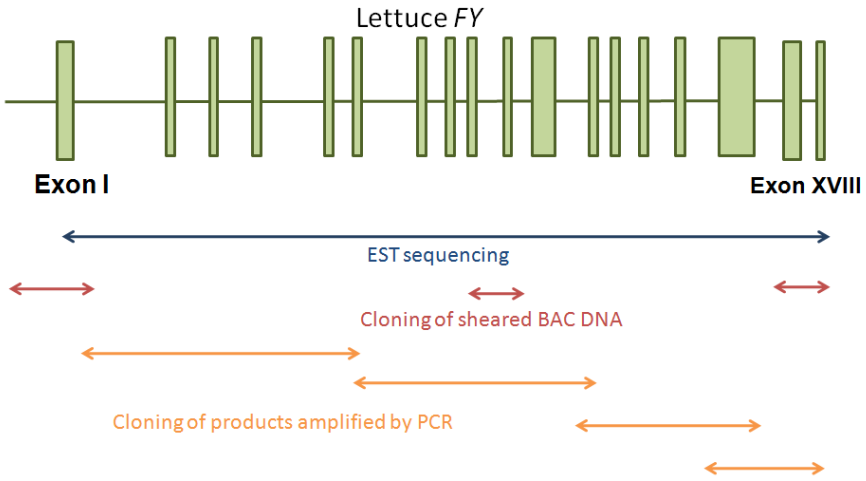


Figure 3.33 - Methods used to isolate a full length lettuce *FY* genomic sequence

Initially, sub-cloning was carried out using BAC clone 094D1, as the hits observed for this clone were extremely strong. Sub-cloning of two bands from a *HindIII* digest of ~5 kb and ~650 bp in size was attempted but proved unsuccessful.

Cloning of sheared BAC DNA fragments was also attempted; sequence covering 300 bp of 5'UTR and *AtFY* exon 1 to intron 1 was obtained, along with sequence

covering *AtFY* introns 6, 7, 15 and 16. PCR reactions using primers (FY_intron1_F and FY_exon6_R), that were designed based on homology between *AtFY* intron1 and exon 6 and BAC clone 094D1, amplified a 2 kb product from BAC 094D1 which was cloned and sequenced. Further PCR reactions using primers (FY_exon6_F and FY_exon 13_R; FY_exon 13_F and FY_exon 16_R; FY_exon 16_F and FY_exon 18_R), based on homology between EST CLPX15054.b1_K19.ab1 and *AtFY* sequences amplified products of 1.6 kb, 1.6 kb and 1.3 kb respectively from BAC 094D1, see figure 3.34, these products were cloned and sequenced.

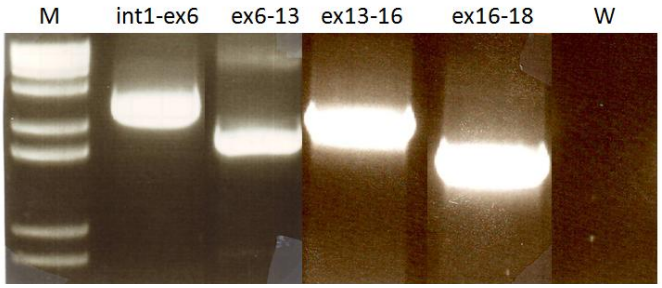


Figure 3.34 – PCR products amplified from BAC clone 094D1. These products were cloned and sequenced. M = 1 kb ladder, W = water control.

Genome walking to obtain more 5'UTR sequence was attempted using BAC 094D1 DNA and genomic DNA as templates. However no more than 485 bp of sequence was available from the BAC clone before running into vector sequence. Although products were amplified from genomic DNA using various pairs of PCR primers, cloning of these fragments always resulted in incorrect sequence.

*Ls*FY is encoded by 665 amino acids comprising 18 exons compared with 647 amino acids which are also encoded by 18 exons in *Arabidopsis*, see figure 3.35. The level of homology between the two species is well conserved; 69.5 % identity, this increases to 84.5 % identity within the WD-40 domain. The top hits obtained from BLASTing *Ls*FY protein are putative WD-40 proteins in grape, castor oil plant, *Medicago* and rice. The top *Arabidopsis* BLAST hit is FY.

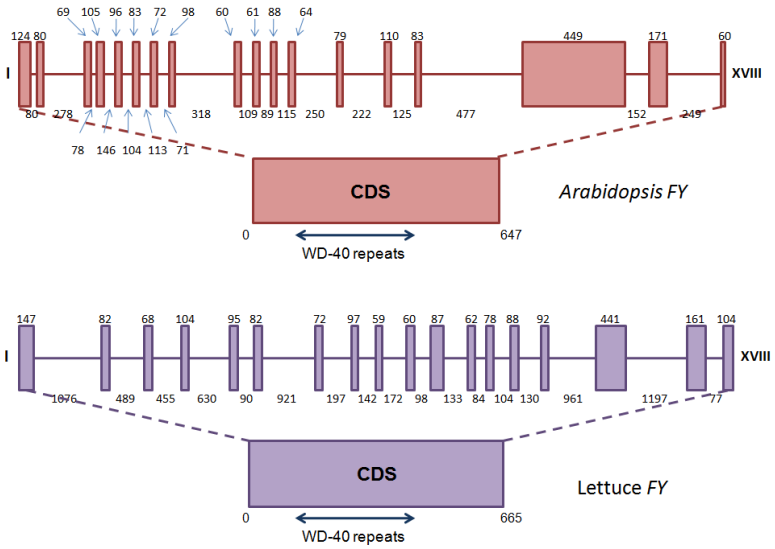


Figure 3.35 - Gene structure *AtFY* and a lettuce FY gene homologue, *Ls*FY. *At*FY is encoded by 647 amino acids arranged in 18 exons; *Ls*FY is encoded by 665 residues also in 18 exons.

3.3.8 FCA

3.3.8.1 Introduction

*At*FCA was identified as a member of the autonomous pathway through the characterisation of *fca-4* (Macknight *et al*., 1997) and *fca-1* (Macknight *et al*., 2002). Both mutations resulted in a delay in flowering; this delay could be restored by subjecting the plants to a vernalisation treatment. *At*FCA contains two RNA recognition motifs (RRMs), at amino acids 121-198 (RRMI) and 212-288 (RRMII), and a WW protein-interaction domain at amino acids 593-624 (Macknight *et al*., 1997). The *AtFCA* transcript is alternatively spliced; only one form encodes the entire *AtFCA* transcript (Macknight *et al*., 2002). *AtFCA* is a predominantly nuclear protein and has the ability to bind RNA. Macknight *et al*., (1997) first suggested that *AtFCA* functions in the post-transcriptional regulation of transcripts involved in the flowering process. As previously mentioned, in section 3.3.7.1, the WW domain of *At*FCA interacts with the PPLP domain of *At*FY to down-regulate expression of *FLC* (Simpson *et al*., 2003). Later work by Bäurle *et al*., (2007) has shown that both *FCA*

and *FPA* are required for RNA-mediated chromatin silencing of a range of loci in the genome, including female gametophytic development and early embryonic development. Further to this it was shown that *AtFCA* and *AtFPA* require *AtFLD* to repress *AtFLC* as described in section 1.4.4 (Bäurle and Dean, 2008; Manzano et al., 2009).

3.3.8.2 Isolation of a lettuce *FCA* (*LsFCA*)

A contig obtained from transcripts expressed in *L. sativa* cv. Salinas; CLS_M3_Contig8775 displayed 68 % identity at the amino acid level with *At*FCA, the contig included EST CLSM1171.b1_E05.ab1 which exhibited 81.9 % identity with RRM II of *At*FCA. Extra sequence from EST CLSM1171.b1_E05.ab1 was obtained using primers M13_R and FCA_seqv2. The sequence obtained displayed homology to *At*FCA 168-747 and the stop codon. The sequence obtained can be seen in figure 3.36, overall 54 % identity with *At*FCA is observed, although there is limited homology within the non-conserved domains of the gene. EST CLSM1171.b1_E05.ab1 displays homology with the final 31 amino acids of RRMI showing 90.3 % identity to the same region in *At*FCA. There is also a significant amount of homology between the WW domains of both species; 89.3 % identity.

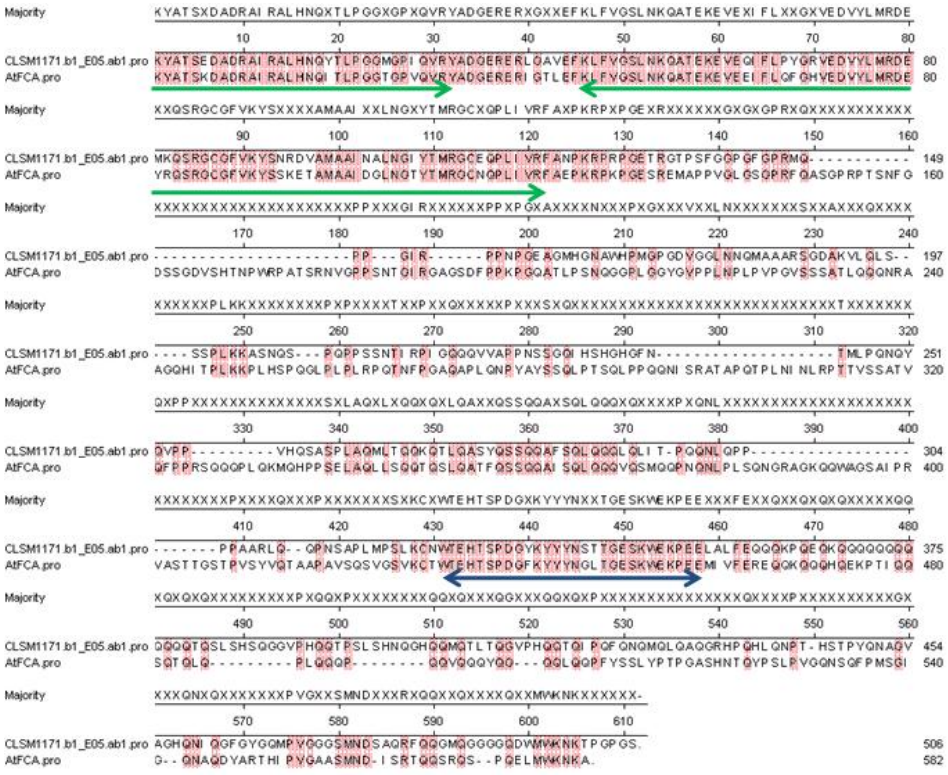


Figure 3.36 – Comparison of EST CLSM1171.b1_E05.ab1 with *At*FCA. RRM domains I and II are highlighted with green arrows and the WW domain is highlighted with a blue arrow.

EST CLSM1171.b1_E05.ab1 was used to screen the lettuce BAC library. 15 strong hits were obtained from the screen. BAC DNA was extracted from each clone and PCR was performed on each sample using primers designed to EST CLSM1171.b1_E05.ab1 (FCA_EST_F1 and FCA_EST_R2). The expected product of 201 bp was amplified from only three of the clones. The other hits observed may have been sequences with homology to the RRM or WW domains. The three BAC clones; 28D12, 57M23 and 20P17, were digested and Southern blotted. Identical banding patterns were seen with all the BAC clones; fragments of 4 kb, 1.8 kb and 1.3 kb were observed when digested with *HindIII* along with fragments of 6 kb and 2 kb with *EcoRV*. Once again, cloning of these products proved to be unsuccessful, so a number of other methods were used to obtain a full length *LsFCA* sequence, see figure 3.37.

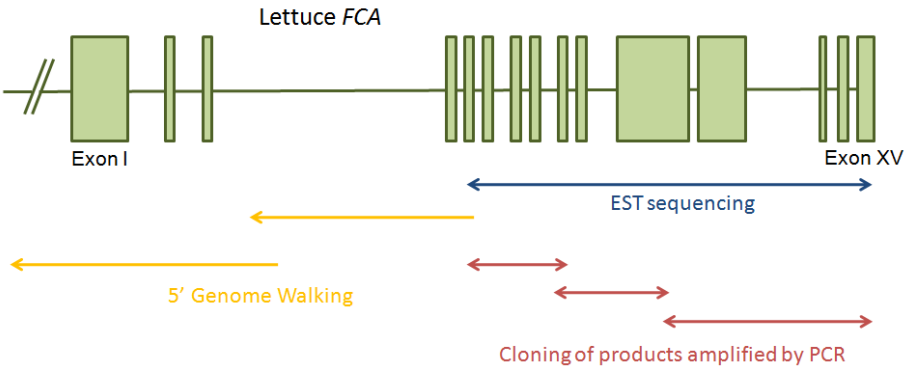


Figure 3.37 - Methods used to isolate a full length lettuce *FCA* genomic sequence

Comparison of the sequence obtained from EST CLSM1171.b1_E05.ab1 and *At*FCA showed that homology between the two species began at *At*FCA exon 5. To obtain further 5' gene sequence genome walking from BAC clone 28D12 was performed. Primers FCA_5_GSP1 and FCA_5_GSP2 amplified a 1.5 kb *EcoRV* PCR product, see figure 3.38, the product was cloned and sequenced, homology was observed with *AtFCA* exon4 and intron3. A second set of genome walk primers were designed to this new sequence; FCA_5_GSP1_2 and FCA_5_GSP2_2 and a 3.5 kb *PvuII* PCR fragment, see figure 3.38, was cloned and sequenced. The product showed homology with the remainder of the coding sequence of *AtFCA* plus ~2 kb of 5'UTR. The genome walk products also covered the first RRM domain of *Ls*FCA; the lettuce sequence displayed 85.9 % identity to *At*FCA in this region.

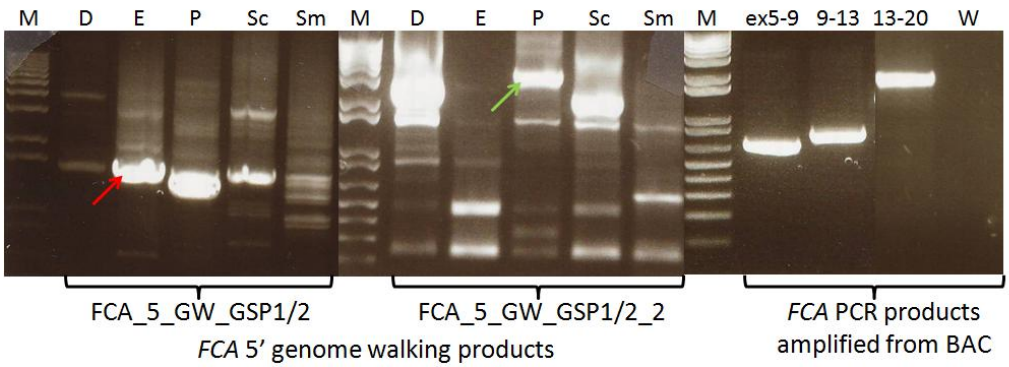


Figure 3.38 – PCR products and genome walk products amplified from BAC clone 28D12, the *EcoRV* and *PvuII* products highlighted with red and green arrows, along with the three PCR products were cloned and sequenced. M = 1 kb+ marker, D = *DraI*, E = *EcoRV*, P = *PvuII*, Sc = *ScaI*, Sm = *SmaI* and W = water control.

The remainder of the gene sequence was obtained by cloning and sequencing PCR products amplified from BAC 28D12. The primers used were designed based on homology existing between EST CLSM1171.b1_E05.ab1 and *AtFCA* exon 5 and exon 9 (FCA_exon5_F and FCA_exon9_R), exon 9 and exon 13 (FCA_exon9_F and FCA_exon13_R), and exon 13 and exon 20 (FCA_exon13_F and FCA_exon20_R). Products of 650 bp, 750 bp and 1.5 kb were amplified, see figure 3.38, cloned and sequenced.

The putative *Ls*FCA gene is encoded by 662 amino acids comprising of 15 exons compared to 20 exons in *At*FCA (747 amino acids), see figure 3.39. *Ls*FCA shows 53.4 % identity with *At*FCA, the highest levels of identity occur within the RRM domains and the WW domain, the remainder of the gene is quite diverged.

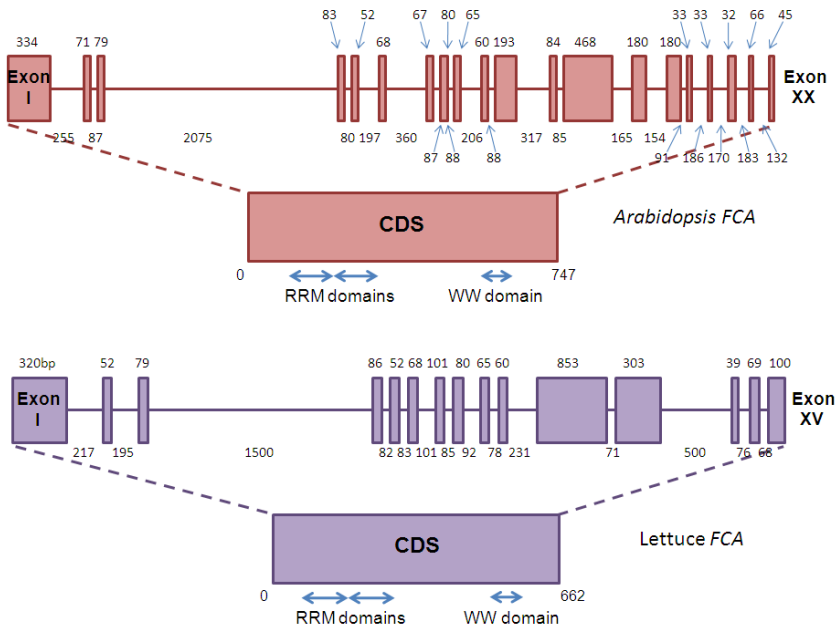


Figure 3.39 - Gene structure *AtFCA* and a lettuce *FCA* gene homologue, *LsFCA*. *At*FCA is encoded by 747 amino acids arranged in 20 exons; *Ls*FY is encoded by 662 residues in 15 exons.

Interestingly rice *FCA* (*OsFCA*), has been characterised by Lee *et al*., (2005), it has 42.9 % identity at the amino acid level with *At*FCA, and 53.8 % with *Ls*FCA. Like *Ls*FCA it is very highly conserved in the RRM and WW domains (78 % and 93 % respectively), but less so in the remainder of the gene. Furthermore the N-terminal

region of *Os*FCA is glycine rich, this is also observed, to a lesser extent, in *Ls*FCA, but not in *At*FCA. A putative pea FCA (*Ps*FCA) is also similar in structure and displays 57 % identity to *Ls*FCA.

3.3.9 FPA

3.3.9.1 Introduction

The *Arabidopsis* gene *FPA* was identified as a member of the autonomous pathway through the characterisation of the alleles *fpa-1*, *fpa-2* and *fpa-6*. These mutants all flower significantly later than WT under both SD and LD photoperiods (Schomburg *et al.*, 2001). The late flowering phenotype is overcome with a vernalisation treatment. *At*FPA protein contains three RRM domains in the N-terminal region, suggesting that it functions as an RNA binding protein (Schomburg *et al*., 2001). *At*FPA is comprised of five exons which encode a protein of 901 amino acid residues. The three RRM domains characterising *At*FPA are located at amino acids 19-87, 96-163 and 207-278, a SPOC domain is also present at amino acids 441-537. *At*FPA acts through the histone demethylase *At*FLD to suppress *AtFLC* expression (Bäurle and Dean, 2008). It is speculated that RNA-binding proteins like *At*FCA and *At*FPA function as part of a transcriptome surveillance mechanism linking RNA recognition with chromatin silencing mechanisms (Bäurle and Dean, 2008). Recent work by Hornyik *et al*., (2010) has identified a role for *At*FPA in controlling the expression of alternatively polyadenylated antisense RNAs at the locus encoding the floral repressor *AtFLC*. This work also suggests that *At*FPA functions redundantly of *At*FCA to control processing of antisense RNAs at the *AtFLC* locus.

3.3.9.2 Isolation of a lettuce *FPA* (*LsFPA*)

Limited homology to *At*FPA was found in the lettuce CPG1 and CPG2 databases. The best hit was a contig with identity at the amino acid level of 47 % over a very small region near the C-terminal end of the gene. The second best hit was a single EST; CLVX2160.b1_O12.ab1, a 186 amino acid transcript expressed in *L. virosa*, which covered *At*FPA amino acid residues 80-276, and had an identity of 40.3 %, but with 72.4 % identity to the final 57 residues of RRM domain III. Further sequence was obtained from EST CLVX2160.b1_O12.ab1 following sequencing with primers M13_R, FPA_seq and FPA_seq2, this sequence showed homology with *At*FPA from amino acid 68-901 and the stop codon. The sequence included the end of RRMII, all of RRMIII, with 69 % identity, and the SPOC domain, with 71.2 % identity, see figure 3.40.

FPA is the top *Arabidopsis* hit when EST CLVX2160.b1_O12.ab1 is BLASTed.

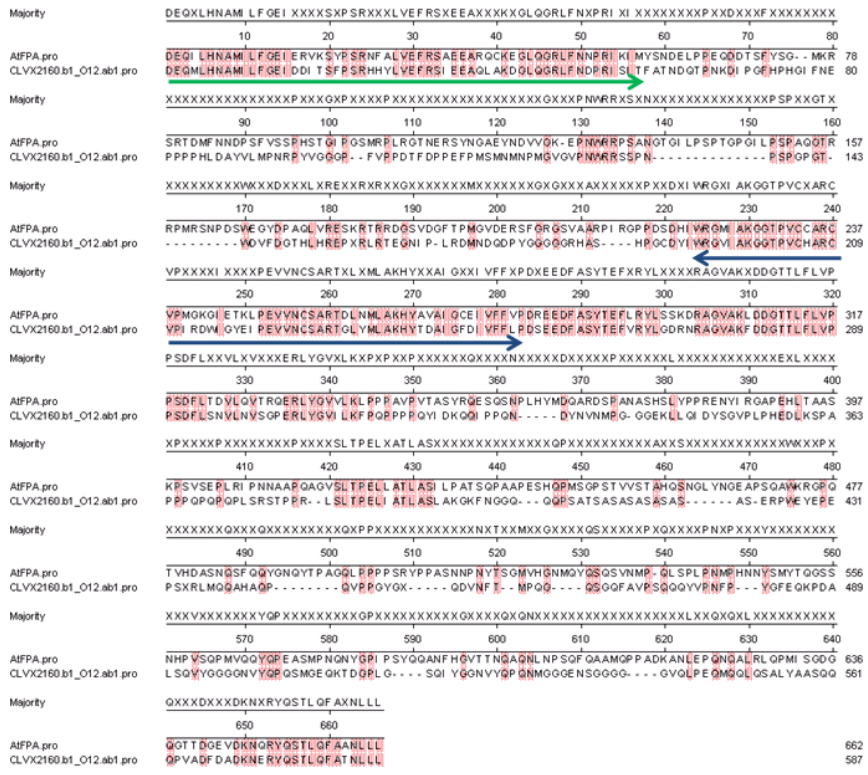


Figure 3.40 – Comparison of EST CLVX2160.b1_O12.ab1 with *At*FPA. RRM domain III is highlighted with a green arrow, the SPOC domain is highlighted with the blue arrow.

EST CLVX2160.b1_O12.ab1 was used to probe the lettuce BAC library. Of the four hits observed only one clone, 041K17 was positively verified by PCR using FPA_EST1_F and FPA_EST1_R primers, designed to EST CLVX2160.b1_O12.ab1, to amplify a 216 bp product. Of the other three hits, only one was a strong hit, suggesting it could be a gene sequence containing RRM domains. BAC DNA from clone 041K19 was extracted and digested and Southern blotted, the blot was probed with EST CLVX2160.b1_O12.ab1, the probe hybridised to three *EcoRV* fragments of 4.5 kb, 2 kb and 1 kb in size, as well as two *HindIII* fragments at 1.6 kb and 1.5 kb. Cloning of these fragments was attempted but was unsuccessful.

A further EST, CLPX10956.b1_G04.ab1 was identified with a high level of identity to the N-terminal end of *At*FPA. This transcript was obtained from *L. perennis* and overlapped EST CLVX2160.b1_O12.ab1 at the 3' end by 96 amino acids with 88 % identity. Figure 3.41 shows the methods used to identify the lettuce *FPA* homologue.

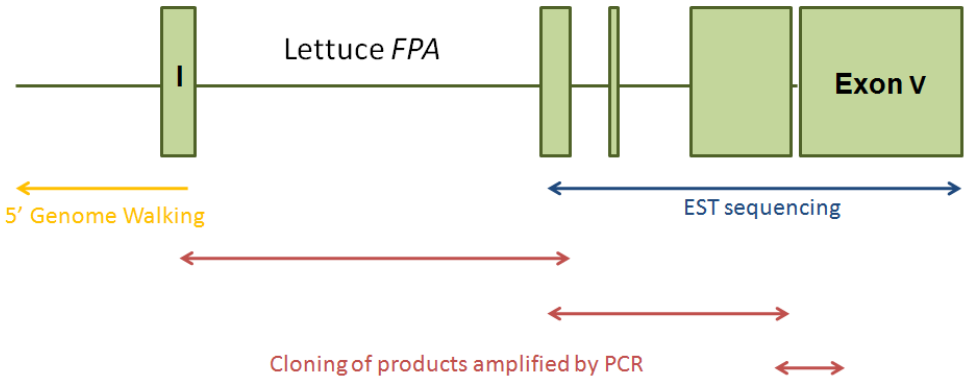


Figure 3.41 - Methods used to isolate a full length lettuce *FPA* genomic sequence

Due to the high level of homology and because EST CLPX10956.b1_G04.ab1 covered RRMI and II domains, a forward PCR primer was designed to this sequence along with a reverse primer designed based on homology between *AtFPA* exon2 and EST CLVX2160.b1_O12.ab1, (FPA_exon1_F and FPA_exon2_R). Further PCR primers were designed based on homology between *AtFPA* and EST CLVX2160.b1_O12.ab1, (FPA_exon2_F and FPA_exon4_R; FPA_exon4_F and

FPA_exon5_R), to obtain full length gene sequence of a putative lettuce *FPA* gene. The 3 kb, 1.8 kb and 250 bp products amplified from BAC clone 041K19, see figure 3.42, were cloned and sequenced. These products produced sequence which identified intronic sequence as well as intron/exon boundaries.

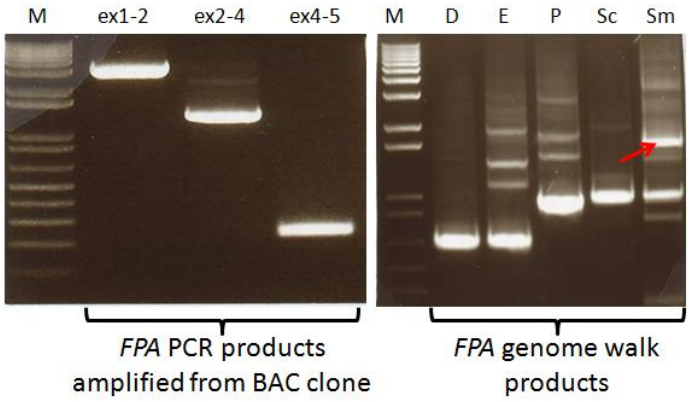


Figure 3.42 – PCR products and genome walk products amplified from BAC clone 041K19, the three PCR products along with the SmaI product highlighted with a red arrow were cloned and sequenced. M = 1 kb+ marker, D = *DraI*, E = *EcoRV*, P = *PvuII*, Sc = *ScaI* and Sm = *SmaI*.

The new sequence obtained from cloning the first FPA_exon1_F – FPA_exon2_R fragment displayed homology to the first two RRM domains, 48.8 % identity with RRMI and 58.8 % identity with RRMII, but not to a potential start codon. To identify the start codon of the *LsFPA* gene and to obtain 5' UTR, genome walking was utilised. Primers FPA_5_GSP1 and FPA_5_GSP2 were designed based on the homology between the sequences obtained from BAC 041K19 and *AtFPA* exon 1. A 1.6 kb product obtained from a *SmaI* genome walk library, see figure 3.42, was cloned and sequenced. A potential start codon and 1.5 kb of 5'UTR was obtained. Figure 3.43, highlights the similarities between *AtFPA* and *LsFPA*. FPA in both species is encoded by five exons, *At*FPA is comprised of 901 amino acids, and *Ls*FPA is slightly shorter at 880 amino acids. Overall, 43 % identity exists between the two species. Only exon 2 is of the same size in both species, *LsFPA* also has a large first intron.

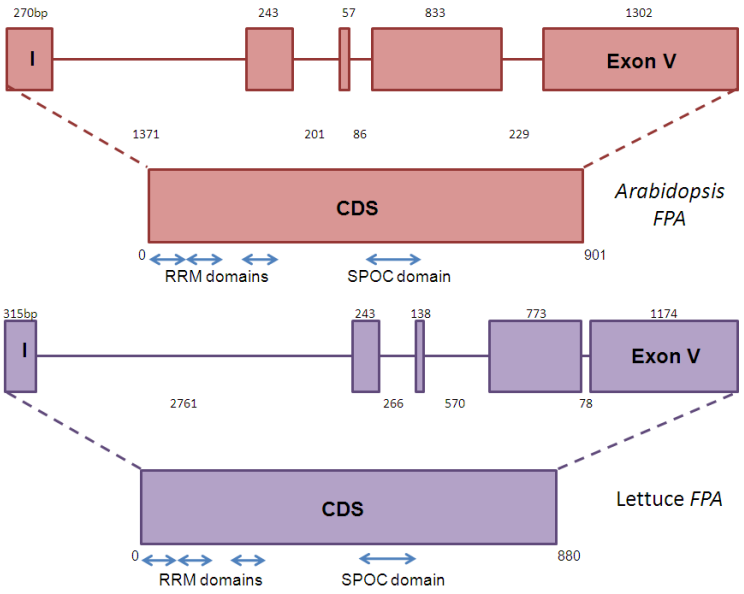


Figure 3.43 - Gene structure *AtFPA* and a lettuce *FPA* gene homologue, *LsFPA*. *At*FPA is encoded by 901 amino acids; *Ls*FPA is encoded by 880 residues, arranged in five exons.

3.3.10 CO

3.3.10.1 Introduction

AtCO is modulated by the circadian clock and day length, (Suárez-López *et al*., 2001), it promotes flowering, by promoting the expression of *AtFT* and *AtSOC1* (Samach *et al*., 2000). As well as being a key gene in the photoperiodic pathway, *AtCO* also makes up part of the light quality pathway.

AtCO belongs to a family of 17 putative transcription factors defined by two conserved domains (Putterill *et al*., 1995). The first is a zinc finger region near the N terminus called the B-boxes, which regulate protein-protein interactions in several animal transcription factors (Borden 1998). The second is a region near the carboxy-terminus termed the CCT domain (Strayer *et al*., 2000), see figure 3.44. *At*CO is comprised of two exons and encodes a protein of 373 amino acids.

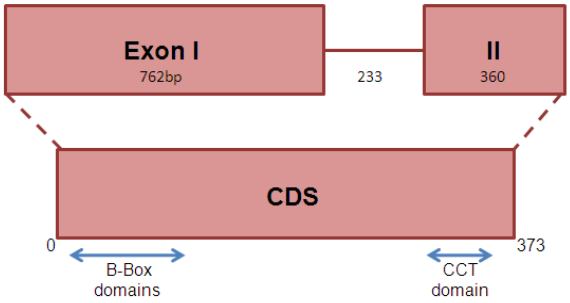


Figure 3.44 – *AtCO* gene structure

Robson *et al*., (2001) showed that the gene family comprising *AtCO* and 16 *Arabidopsis CO-like (AtCOL)* is subdivided into three groupings depending on the number and type of B-boxes present. The phylogenetic tree in figure 3.45 highlights the three groups, the first contains *AtCO* and *AtCOLs1-5* that contain two B-boxes, the second contains *AtCOLs6-8* and *AtCOL16* which each have only one B-box. The third group is comprised of *AtCOLs9-15*, each contain two B-boxes, the second of which is diverged from those genes making up Group I.

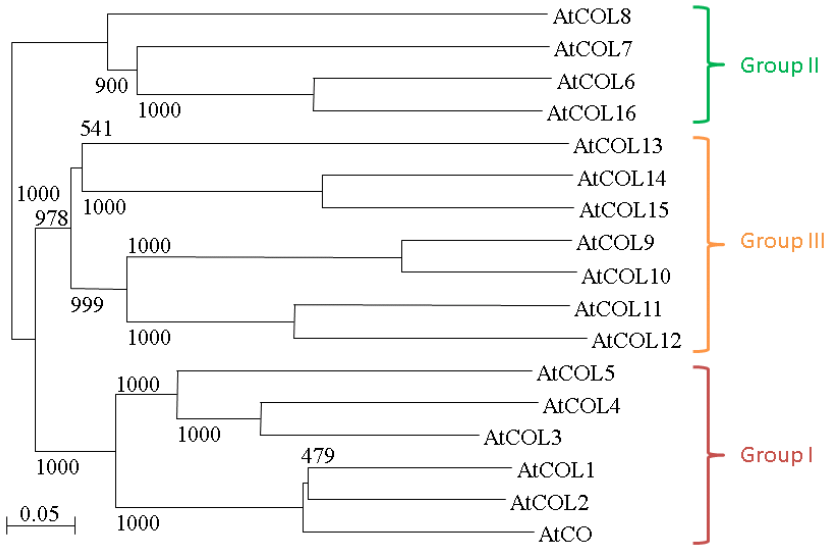


Figure 3.45 – Phylogeny of the *Arabidopsis CO* and *COL* genes. The 17 genes can be placed in one of three groups based on the number and type of B-box present.

3.3.10.2 Analysis of a lettuce EST contig with homology to *AtCO*

Most of the hits obtained from the lettuce database displayed high levels of homology with the CCT domain of *AtCO*, this domain is highly conserved in *AtCO* and the *AtCOL* genes and is also present in a number of other *Arabidopsis* genes, for example *TOC1, ZML1 (ZIM-LIKE 1),* and the *APRR (ARABIDOPSIS PSEUDO-RESPONSE REGULATOR)* gene family. QG_CA_Contig3669, created from transcripts expressed in *L. sativa* cv. Salinas, with 34.1 % identity to *At*CO at the amino acid level, contained the EST QGF10H03.yg.ab1 which displayed a 38.1 % identity to *At*CO at the amino acid level. Interestingly this was one of a few ESTs to show any homology to both the B-boxes, suggesting that the EST could represent a lettuce *CO* or *COL1-5* gene homologue. A level of 58 % identity with *At*CO was observed within the B-box regions, see figure 3.46 for details. Because both B-boxes are represented in EST QGF10H03.yg.ab1, it was selected as an EST to follow up. QG_CA_Contig3669 also contained the EST QGJ4D05.yg.ab1 which covered the CCT domain of the *At*CO/*At*COL proteins.

	*At*CO	*At*COL1	*At*COL2	*At*COL3	*At*COL4	*At*COL5
QGF10H03.yg.ab 1 (B-box region)	58%	67.4%	62.8%	77.9%	81.4%	69.5%
QGJ4D05.yg.ab1 (CCT domain)	83.7%	83.7%	83.7%	90.7%	88.4%	90.7%

Figure 3.46 – Comparison of *At*CO and COL1-5 B-box regions with EST QGF10H03.yg.ab1, and the CCT domain with EST QGJ4D05.yg.ab1. Green and blue arrows highlight B-box 1 and 2 respectively and the orange arrow highlights the CCT domain in *Arabidopsis CO* and *COL1-5* genes. The table illustrates the level of identity at the amino acid level between lettuce ESTs QGF10H03.yg.ab1 and QGJ4D05.yg.ab1with *At*CO and *At*COL1-5 B-boxes and the CCT domain respectively

A forward PCR primer was designed to the second B-box of EST QGF10H03.yg.ab1 and a reverse primer was designed to the highly conserved CCT domain sequence in EST QGJ4D05.yg.ab1. An 822 bp PCR product was amplified and sequenced; the sequence obtained matched that of the ESTs making up QG_CA_Contig3669 and produced sequence corresponding to the middle region of the gene. The amino acid sequence of this fragment displayed 38.8 % identity with *At*CO; but showed a higher identity to COL3 (51 %), COL4 (53.5 %) and COL5 (46.2 %). Phylogenetic analysis of QG_CA_Contig3669 confirms that it most closely resembles *At*COL4 than *At*CO, see figure 3.47.

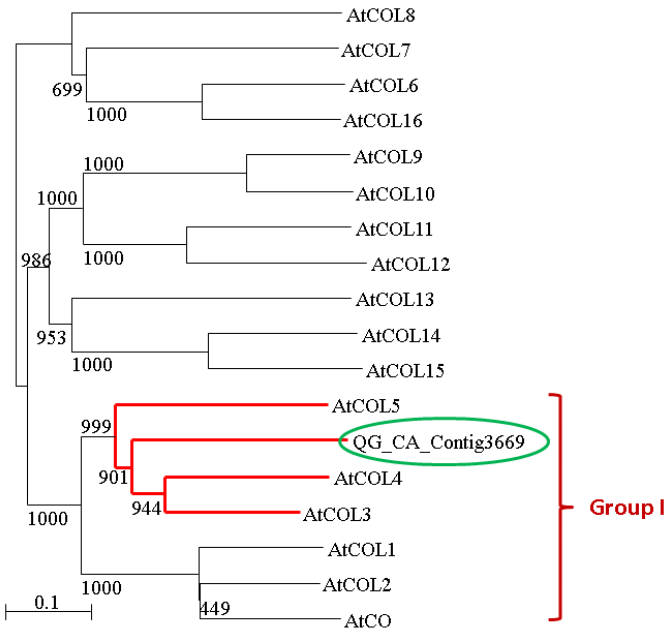


Figure 3.47 – Phylogenetic tree illustrating the position of lettuce EST QG_CA_Contig3669 compared to the *Arabidopsis CO* and *COL* family genes. The EST shows most homology with *At*COL4.

EST QGF10H03 was selected to probe the lettuce BAC library with the aim of obtaining a functional lettuce *CO*. Five hits were observed on probing the BAC library, DNA was extracted from these clones and PCR verification was carried out using primers designed to the CCT domain (CO_CCT_2F and CO_CCT _R) and primers designed from the second B-box to the CCT domain, (CO_B-BOX-CCT_F1 and CO_B-BOX-CCT_R1). Four of the clones; 176G13, 11I13, 28M04 and 30B23 produced the expected 111 bp and 740 bp products. However only the 111 bp product from the CCT domain was amplified from BAC clone 8B14, suggesting that there is no second B-box in this BAC clone, therefore it was most likely to contain a sequence with homology to an *AtCOL* gene from groups II or III. The PCR fragments amplified from each of the BAC clones were sequenced, all were identical to one another and to that of QG_CA_Contig3669, with highest homology to *AtCOL4*. To further verify the clones obtained, BAC DNA was extracted, digested and Southern blotted, EST QGF10H03 was used as a probe and it hybridised to

fragments of identical sizes in each of the clones; an *EcoRV* product of 4.5 kb and a *DraI* product of 3.7 kb.

As *AtCO* has a very distinctive expression profile which is different to that of *AtCOL1* and *AtCOL2* genes (Ledger *et al*., 2001), quantitative real time PCR was used to look at the expression profile of the putative lettuce *CO* sequence obtained. It is known that *AtCO* displays clear expression peaks at approximately 16h after dawn in LD photoperiods (Suárez-López *et al*., 2001). The expression peak of *CO* in *Arabidopsis* is crucial to the gene's function within the photoperiod pathway as described in section 1.42. The expression of *AtCOL1* and *AtCOL2* is very different, they show maximum expression at the beginning of the day (Ledger *et al*., 2001). cDNA was synthesised from lettuce leaf material collected every 2h over a 24h period, primers specific to *LsCOL* homologue were used to measure its expression over this time course and the data obtained was normalised to the expression of *LsELFα*. Expression of *LsCOL* increases after 12h rising to a peak in the dark 20h after subjective dawn, see figure 3.48. This data suggests that the lettuce gene expression pattern is more similar to *AtCO* than to *AtCOL1* or *AtCOL2* which are also Group I *Arabidopsis COL* genes. The phylogenetic analysis suggests that the *LsCOL* is most like *AtCOL4*, the expression profile for *AtCOL4* has not been analysed, and neither has that of the other Group 1 genes, *AtCOL3* and *AtCOL5*.

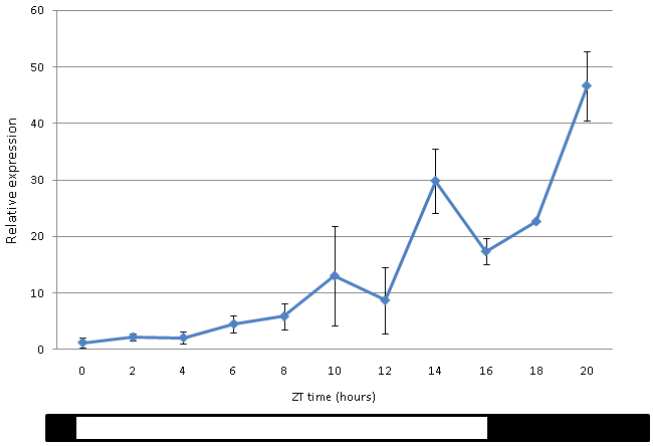


Figure 3.48 – The expression of *LsCOL* over a 24h period, relative to *LsELFα*. The white/black bars denote light/dark cycles.

The identification of a *CO* orthologue in crop species has been shown to be notoriously difficult (Hecht *et al*., 2005; Taylor, 2009). Because all the *CO* sequences obtained from the lettuce BAC library show highest homology to *AtCOL4*, and because of the amount of work that would need to be undertaken to identify a lettuce *CO* orthologue, it was decided not to follow up *CO* as a target gene in this project.

It would be interesting to attempt to complement the *Arabidopsis co-2* mutant with the *LsCOL* gene homologue. It has been shown that an altered circadian expression of a functional CO is seen in barley; this suggests that differences in the expression of conserved genes vary between species.

3.3.11 CRY2

3.3.11.1 Introduction

There are nine photosensory receptors in *Arabidopsis*, including the phytochromes, (*PHYA-E*) and the cryptochromes (*CRY1* and *CRY2*) which are involved in the regulation of flowering. *AtCRY2* is involved in mediating light input to the circadian clock (Devlin and Kay, 2000), and its expression changes in response to photoperiod

(El-Assal *et al*., 2003). *AtCRY2* is the major blue light photoreceptor affecting flowering, CRY2 protein together with PHYA, acts to stabilise the CO protein towards the end of a LD (Valverde *et al*., 2004). A further role for *AtCRY2* has been identified in the regulation of *AtFT* expression in vascular bundles, possibly through stabilisation of the CO protein (Endo *et al*., 2007).

AtCRY2 is comprised of four exons encoded by 612 amino acids. *AtCRY2* is a nuclear protein which contains a DNA photolyase domain at amino acids 7-172, and a Flavin-adenine dinucleotide (FAD) binding 7 domain (Cashmore *et al*., 1999), at amino acids 212-489, see figure 3.49.

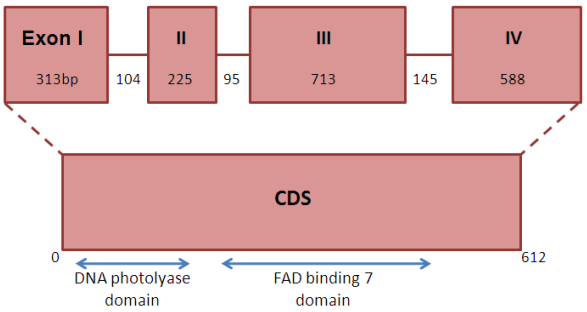


Figure 3.49 – Gene structure of *AtCRY2*

3.3.11.2 Analysis of a lettuce EST with homology to *CRY2 (LsCRY2*)

EST CLSM20253.b1_I24, a transcript expressed in *L. sativa* cv. Salinas displayed homology to the DNA photolyase domain of *At*CRY2. Further sequence was obtained from EST CLSM20253.b1_I24 using M13_R, CRY2_exon3_R and CRY2_ESTseq primers. This sequence covered the first 416 amino acids of *At*CRY2 and exhibited 68 % identity, it showed just 57 % identity with *At*CRY1. The sequence covered the DNA photolyase domain and showed 68.1 % identity, it also covered first 205 amino acids of the FAD binding 7 domain, displaying 71.2 % identity in this region, see figure 3.50

Figure 3.50 – Comparison of *At*CRY2 and EST CLSM20253.b1_I24. The DNA photolyase domain is highlighted with a green arrow and the FAD binding 7 domain is highlighted with a blue arrow.

EST CLSM20253.b1_I24 was used to probe the lettuce BAC library, four hits were observed; BAC DNA was extracted from each of the four clones and was verified by PCR. Primers designed specifically to EST CLSM20253.b1_I24, CRY2_exon3_F and CRY2_exon3_R amplified a PCR product of 130 bp, from each clone which were sequenced verified. BAC clone 120A17 produced a different sequence to the lettuce EST and the other BAC clones, this sequence displayed highest levels of homology with *AtCRY1*. BAC clone 120A17 was therefore discounted at this stage. Genomic DNA was extracted from the other three BAC clones; 37H05, 163J11 and 29I08, the DNA was digested and Southern blotted. EST CLSM20253.b1_I24 was used as a probe and hybridised to fragments of 3 kb and 1.5 kb from an *SspI* digest. It was decided that a full length *CRY2* gene sequence would not be required for this project because mutations affecting *CRY2* function could be detected using a phenotypic screen of the LB lettuce lines. It has been shown in *Arabidopsis* that plants containing mutations in *CRY2* produce longer hypocotyls than plants not containing a mutation under continuous blue light. It was decided that this method

could be used to screen the LB lettuce lines identified for mutations within *LsCRY2*, see sections 5.2.17 and 5.3.1.5.

3.3.12 FKF1

3.3.12.1 Introduction

AtFKF1 is a circadian-regulated gene which affects flowering time. *At*FKF1 protein has three characteristic domains: an N-terminal PAS motif, similar to the light, oxygen and voltage (LOV)-sensing domain of certain photoreceptors, an F-box which is found in proteins involved in targeting ubiquitin-mediated degradation, and six kelch repeat motifs predicted to be involved in protein-protein interactions (Nelson *et al*., 2000; Nagasako *et al*., 2005), see figure 3.51. It has been shown to function as a photoreceptor for the light controlled expression of *CO* (Nelson *et al*., 2000). FKF1 protein also mediates the cyclic degradation of *CYCLING DOF FACTOR 1* (*CDF1*), a repressor of *CO* (Imaizumi *et al*., 2005). *At*FKF1 is closely related to ZEITLUPE, (ZTL), displaying 62 % identity at the amino acid level.

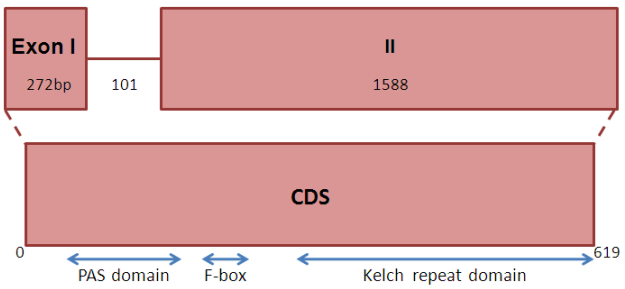


Figure 3.51 – Gene structure of *AtFKF1*

3.3.12.2 Analysis of a lettuce EST with homology to *FKF1 (LsFKF1*)

EST QGG32A02.yg.ab1, a transcript expressed in *L. sativa* cv. Salinas displayed 63.5 % identity with *At*FKF1 at the amino acid level, and covered the C-terminal end of the PAS domain (84.4 % identity), the F-box (84.8 % identity), and the first 29

residues of the Kelch repeats. The EST was sequenced using M13_R and FKF1_ESTseq primers and an extra 301 amino acid residues were obtained, covering the remainder of *At*FKF1, including the Kelch repeat domain (80.9 % identity). Overall EST QGG32A02.yg.ab1 displays 78.6 % identity with *At*FKF1 and covers the protein from amino acid 135 to the stop codon, see figure 3.52. EST QGG32A02.yg.ab1 also shows 64 % identity with *At*ZTL.

Figure 3.52 - Comparison of *At*FKF1 and EST QGG32A02.yg.ab1. The C-terminal end of the PAS domain is highlighted with a green arrow, the F-box is highlighted with a blue arrow and the kelch repeat domain is highlighted with a red arrow.

EST QGG32A02.yg.ab1 was selected to probe the lettuce BAC library. Three hits were observed; the clones were identified and were PCR verified using FKF1_EST_F and FKF1_EST_R primers which had been designed specifically to EST QGG32A02.yg.ab1. The expected PCR product of 260 bp was amplified from only one BAC; clone 158H06. This was the strongest hit observed from the library screen. DNA was extracted from the clone and was digested and Southern blotted, EST QGG32A02.yg.ab1 was used as a probe. A 2.7 kb fragment from an *EcoRV* digest and a 5 kb *SmaI* generated fragment both hybridised to the probe. Cloning of both products was attempted unsuccessfully. Work on identifying a lettuce *FKF1*

homologue stopped at this point for this project. The work was continued by Dr. Andrea Massiah, Mark Kerr and Jemma Taylor, WHRI. They genome walked from BAC clone 158H06 and obtained the remainder of the 5' coding sequence, including the remainder of the PAS domain (89.4 % identity). The full length *Ls*FKF1 coding sequence was 640 amino acid residues in length and arranged in 3 exons, compared to 619 residues in 2 exons in *Arabidopsis*. The two genes share 76.8 % identity, (*Ls*FKF1 shows 63.1 % identity with *At*ZTL).

3.3.13 FLC

3.3.13.1 Introduction

AtFLC is a key floral repressor of flowering (Ratcliffe *et al*., 2003). It is the convergence point of the vernalisation and autonomous pathways. *AtFLC* expression is repressed by vernalisation (Michaels and Amasino, 1999), and is also promoted by the genes making up autonomous pathway. *At*FLC is closely related to the five MADS AFFECTING FLOWERING (MAF) proteins which display 53-87 % identity, see figure 3.53 (Bodt *et al*., 2003; Ratcliffe *et al*., 2003).

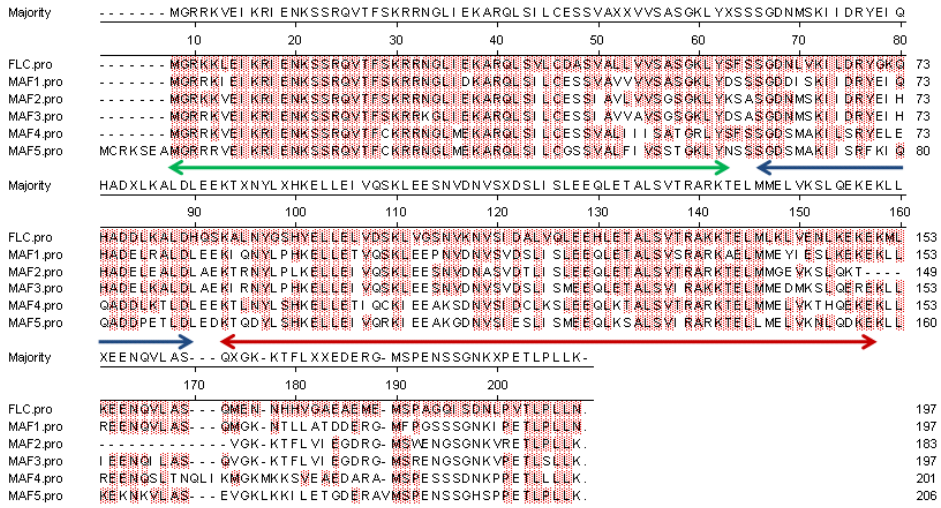


Figure 3.53 – Sequence comparison of *At*FLC and *At*MAF1-5. *MAF1-5* are *FLC* paralog genes in *Arabidopsis*, all contain a MADS box, highlighted with a green arrow, an I-Box, highlighted with a blue arrow and a K-Box region, highlighted with a red arrow.

MAF1, also known as FLOWERING LOCUS M (FLM), like *At*FLC is a repressor of flowering (Ratcliffe *et al*., 2001; Scortecci *et al*., 2001). Ratcliffe *et al*., (2001), also suggest that MAF2-4 act like *At*FLC and repress flowering, they also suggested that MAF5 may act as a floral promoter. It was also shown that vernalisation represses *MAF1-3*, but induces *MAF5* and does not have a strong affect on *MAF4* (Ratcliffe *et al*., 2003; Sung *et al*., 2006).

*At*FLC is composed of seven exons, the first of which contains a MADS box domain, the gene also contains both I box and K box domains, see figure 3.54, these domains are well conserved between *At*FLC and the MAF family genes.

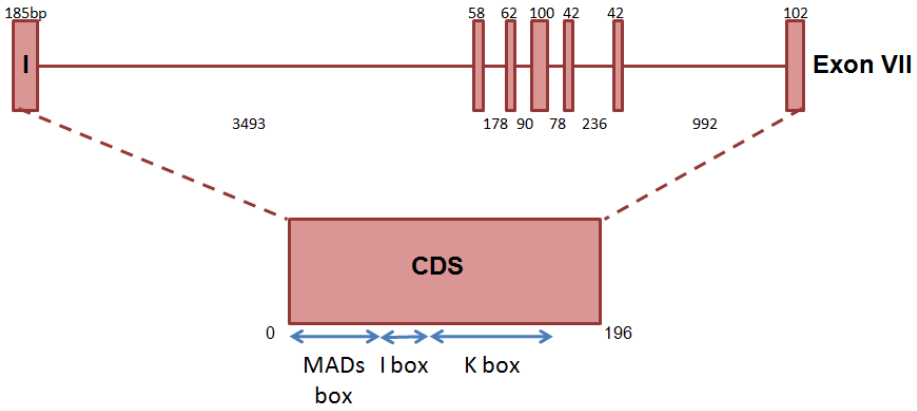


Figure 3.54 – Gene structure of *AtFLC*

3.3.13.2 Analysis of ESTs with homology to *AtFLC*

Many ESTs showed homology to *AtFLC*, the best of which was CLS_S3_Contig593 (displaying 45.5 % identity), however the similarity observed for each of these ESTs occurred within the MADS box domain, a region which is present in a large number of genes in *Arabidopsis*. A BLAST search using the *At*FLC MADS box domain returns over 100 hits, because of this non-specificity it was decided that using any EST containing the MADS box domain for further analysis would be difficult, for example using such an EST would result in multiple BAC library hits.

A PCR approach was therefore undertaken; primers were designed based on the *AtFLC* sequence. *AtFLC* and *AtMAF1-5* family genes are highly conserved so primers were designed to the most divergent regions between the gene sequences. PCR primers were designed specifically to the I-Box and K-Box regions of *AtFLC*, ensuring the primers were significantly different from the same region in the other MAF family genes. No products were amplified from lettuce, however a product of 421 bp was amplified from *Arabidopsis* ecotype Col-0 genomic DNA this product was sequence verified and was used to probe the lettuce BAC library. Six hits were observed from the BAC library screen, however when attempting to verify these clones using PCR no products were amplified using three sets of primers designed to *AtFLC*.

A Southern blot was constructed using lettuce genomic DNA digested with *BamHI* and *EcoRV*, the blot was probed with the 421 bp PCR products described above. The probe hybridised to two *BamHI* products of 2.6 kb and 1.4 kb in size. Cloning of these products was attempted without success.

As with *CO*, a functional *FLC* orthologue has proved extremely difficult to obtain from other plant species. It was decided that no further work would take place on this gene in this project. As with *FKF1*, work by Dr. Andrea Massiah, Mark Kerr and Jemma Taylor based on Contig CLS_S3_Contig593, plus three ESTs identified from other databases has identified coding sequence corresponding to four lettuce *FLC* like genes (*LsFLC* genes 1-4); gene 1 displays the highest level of identity with *At*FLC at the amino acid level; 68.2 %.

3.3.14 Summary

Of the twelve *Arabidopsis* flowering time genes targeted, full length genomic sequences of nine have been identified in lettuce. The details of each gene homologue are summarised in table 3.6. A number of ESTs, from wild lettuce as well as a cultivated variety of lettuce, were identified as showing homology to each of the genes targeted. Using sequence comparisons and by constructing phylogenetic trees as well as identifying conserved sequence domains and taking into consideration whether the gene was part of a family, the best EST representing each of the *Arabidopsis* flowering genes in lettuce was identified. The ESTs were used to probe a lettuce BAC DNA library, BAC DNA clones from positive hits were verified before being digested with a number of restriction enzymes and Southern blotted. Interestingly, all the BAC clones obtained from each hybridisation produced the same banding patterns on the Southern blot. This suggests that from each screen the BAC clones identified probably cover the same region of the lettuce genome, unless there are very large regions of duplication in the lettuce genome. This approach meant that the search for the lettuce target genes was limited by using a single EST sequence for each target gene, as opposed to following up multiple EST sequences. Although subcloning of the digested BAC DNA fragments containing the lettuce homologue sequences was largely unsuccessful, other methods were also utilised to obtain the full length sequences, e.g. PCR, direct sequencing of the BAC DNA, and genome walking from the BAC DNA using primers designed to the EST sequences.

The level of identity at the amino acid level between the *Arabidopsis* flowering gene and the homologue identified in lettuce ranges from 43-79 % (FPA and FVE). All nine lettuce homologues isolated contain specific domains which have been used to characterise the gene in *Arabidopsis*. Six of the gene homologues identified in

lettuce contain the same number of exons as in *Arabidopsis*; *FT*, *FLK*, *LD*, *FVE*, *FPA* and *FY*. Only one gene *FT*, is encoded by an identical number of residues in each species; 176 amino acids.

	No. ESTs/Contigs with homology	No. BAC library hits	No. BAC clones PCR verified	Protein Size (aa)		Identity (%)
				Arabidopsis	Lettuce	
FT	1	1	1	176	176	75
FLK	24	3	2	577	485	55.6
FLD	6	6	3	789	848	69.2
LD	1	1	1	953	960	47.2
FVE	18	7	3	507	460	79
FY	24	4	2	647	665	69.5
FCA	25	15	3	747	662	53.4
FPA	10	4	1	901	880	43
CO	23	5	4	373	N/A	N/A
CRY2	16	4	3	612	N/A	N/A
FKF1	18	3	1	619	640	76.58
FLC	24	6	0	196	N/A	N/A

Table 3.7 – Summary of *Arabidopsis* flowering time genes targeted in lettuce. The table details the number of lettuce EST/contigs analysed, and the number of BAC hits recorded upon probing a lettuce BAC library with the best EST clone. The number of the BAC clones that were positively verified is also included along with a comparison of the size of the gene homologues identified and the level of identity between the species at the amino acid level.

Of the remaining eight genes, four are larger in size in *Arabidopsis* and four are larger in lettuce. Three *Arabidopsis* genes are significantly larger than their lettuce counterparts; *AtFLK* and *AtFVE* both have a larger exon 1 than the lettuce homologues of these genes, whereas the difference in size in *AtFCA* seems to be due to there being more exons present than in lettuce, 20 compared to 15. Only one lettuce gene is significantly larger than in *Arabidopsis*, *LsFLD* is encoded by one large exon as opposed to the two making up the coding sequence of *AtFLD*.

Perhaps of most interest are the *FT* and *LD* lettuce gene homologues identified, both genes were represented by only one EST in the CGP database, and only one clone with homology was identified upon screening a lettuce BAC DNA library. The ability of *LsFT* to complement the function of *AtFT* in restoring the flowering

phenotype of the *Arabidopsis* mutant *ft-1* is evidence that *LsFT* is a functional *AtFT* orthologue. It has also been shown that *LsFLK* complements the function of *AtFLK* and restores the flowering phenotype of the *flk-4* mutant in *Arabidopsis*, providing evidence that *LsFLK* is a functional gene orthologue of *AtFLK*. It is possible that *LsFT* and *AtFLK* are not the only functional *AtFT/AtFLK* gene orthologues in lettuce; a Southern blot to look at the copy number of *LsFT/LsFLK* within the lettuce genome would be a future experiment to carry out.

Interestingly a number of the lettuce genes identified show a higher level of homology at the amino acid level to putative flowering time genes in other crop species than to the *Arabidopsis* gene, this provides confidence that the lettuce genes isolated are good flowering time gene homologues. *LsFT* is more highly conserved to *FT* homologues in sunflower and grape, *LsFLK* to *FLK* homologues in grape and rice, *LsLD* and *LsFY* to grape, *LsFY* also to *Medicago* and rice and *LsFCA* to rice and pea.

The lettuce flowering gene sequences that have been obtained will be analysed in all those lettuce lines identified as having a LB phenotype to look for polymorphisms that may be affecting the function of those genes and thus causing the LB phenotype. At least 1 kb of 5' untranslated region was identified from all the genes with the exception of *LsFY*. This means that mutations within promoter sequences that may affect the expression of the flowering genes can also be identified. The full length lettuce gene sequence obtained for *FT* and the seven autonomous pathway genes can be found in Appendix VI.

CHAPTER 4

Identification of late bolting lettuce lines

4.1 Introduction

4.1.1 Overview

This section of the project involved identifying LB lettuce lines by screening populations containing induced mutations and looking at naturally occurring allelic diversity in wild lettuce species, see figure 4.1.

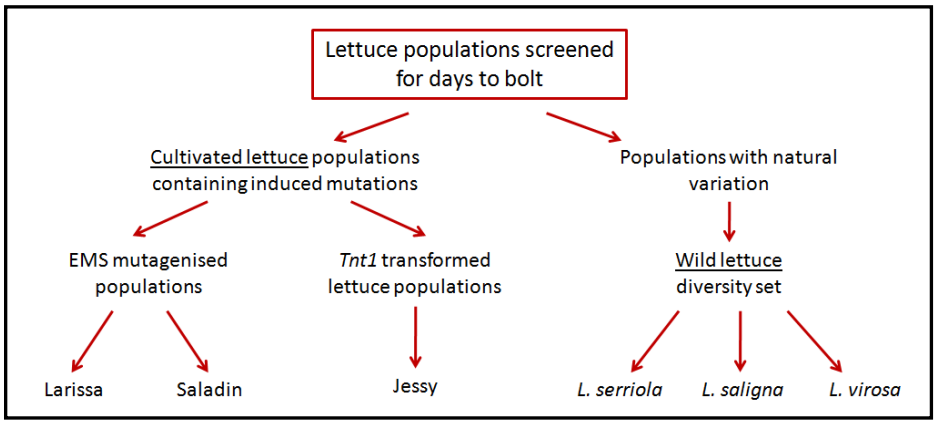


Figure 4.1 – Details of the lettuce populations screened for bolting time. Two lettuce cultivars (Larissa and Saladin) were treated with EMS to induce mutations throughout the populations; a further cultivar (Jessy) was transformed with the tobacco *Tnt1* retrotransposon element. Populations of three wild lettuce species (*L. serriola*, *L. saligna* and *L. virosa*) were also screened to identify natural genetic variation affecting bolting time.

The populations containing induced mutations were made up of two EMS mutagenised populations in the cultivars *L. sativa* cv. Larissa and *L. sativa* cv. Saladin, (butterhead and iceberg types of lettuce respectively). A small lettuce population, *L. sativa* cv. Jessy (a butterhead style lettuce), was also screened, this population had been transformed with the tobacco *Tnt1* retrotransposon element (Mazier *et al.*, 2007) and is thus classed as GM. Because of the agronomic implications of this project the use of GM plants was kept to a minimum and these *Tnt1* lines were used purely for research purposes. The final population that was screened was a wild lettuce diversity set to look for alleles causing natural variation

in bolting time, the set was comprised of lines from three species, *L. serriola*, *L. saligna* and *L. virosa.*

4.1.2 Strategy

The Larissa and Saladin EMS populations were the main focus of the research. The Larissa population, which was a test population for the EMS mutagenesis protocol, was relatively small and so many replicates could be initially screened in the glasshouse. Larissa has a relatively short life cycle, and therefore multiple generations of the plants could be grown each year. With this in mind it was important to confirm the phenotype of the interesting lines under different photoperiods. EB lines were also initially identified as important, with the intention of using these plants to aid the search for flowering time genes, any mutations located in these lines which could be located to a flowering gene homologue would help in the screening process of the LB lines. The majority of the work on the Larissa population took place in the glasshouse under controlled temperatures and the plants were randomised using alpha block designs (Mead *et al*., 1993), throughout to ensure the replication of any interesting phenotype observed.

Plants were deemed to have initiated bolting when floral parts could first be identified, see figure 4.2.

Figure 4.2 – Criteria for scoring lettuce plants - lettuce plants were scored as bolted when floral parts became visible as indicated with the red arrow. This example is a *L. sativa* cv. Larissa plant.

Plants were scored for number of days to bolt from being sown. The number of leaves, another commonly used method for scoring bolting or flowering, was not employed as the number of leaves produced by lettuce is difficult to ascertain and leaves were removed as the plant got older; too many leaves can prevent lettuce head development and can cause problems in the development of floral parts, also leaves in direct contact with the bench/soil can encourage the spread of disease. The number of days to the first open flower was also scored for some of the experiments, to ascertain whether the time of bolting is directly related to the time of flowering.

To remove some of the background EMS induced mutations the interesting lines were BC to WT twice before being selfed. This enabled the identification of homozygous LB mutant lines; the robustness of the phenotype was tested with and without a vernalisation treatment to look for the presence of a mutation within an autonomous pathway gene. Plants which were subjected to a vernalisation treatment were scored for number of days to bolt by combining the number of days pre- and post- the treatment; the four weeks that the plants were vernalised were not included because very little plant growth occurs at 4 °C.

The LB phenotype was also tested in the field, an important aspect of the work was to ensure the LB phenotype of lines in the field under conditions typical to that of a

lettuce farm, and replication of the field trials enabled the comparison of the plants bolting phenotype under varying photoperiods. The response of the plants was also tested under different temperatures; with the trend for warmer summers with the threat of climate change it was important to test the response of the plants to high temperatures. A similar strategy was applied to the Saladin populations, except due to the number of plants involved all were screened in the field under the protection of plastic haygrove polytunnels. Once a LB line was confirmed in the field the line was re-scored in the glasshouse, if the LB phenotype was confirmed BC commenced as with the Larissa population. Only one generation of Saladin could be grown per year as this cultivar had a light quality requirement which could not be satisfied by a SD photoperiod or the supplementary lighting available. This meant that not as much progress could be made with the Saladin population as with the Larissa population. All experiments relating to the Jessy population and the wild diversity set were screened in the glasshouse only.

4.2 General materials and methods

4.2.1 Field trial conditions

Plants for field trials were sown and transplanted as described in section 2.1.2.1. Plants were transplanted into the field and were protected by plastic poly-tunnel covering, see figure 4.3.

Figure 4.3 – Saladin field trial conditions – Plants raised in modular trays and reared in the glasshouse before being transplanted in field under poly tunnels.

Plants were transplanted in a standard layout as shown in figure 4.4. Plants were not randomised in these trials, due to the large number of plants being used. Plants were watered using a drip feed system.

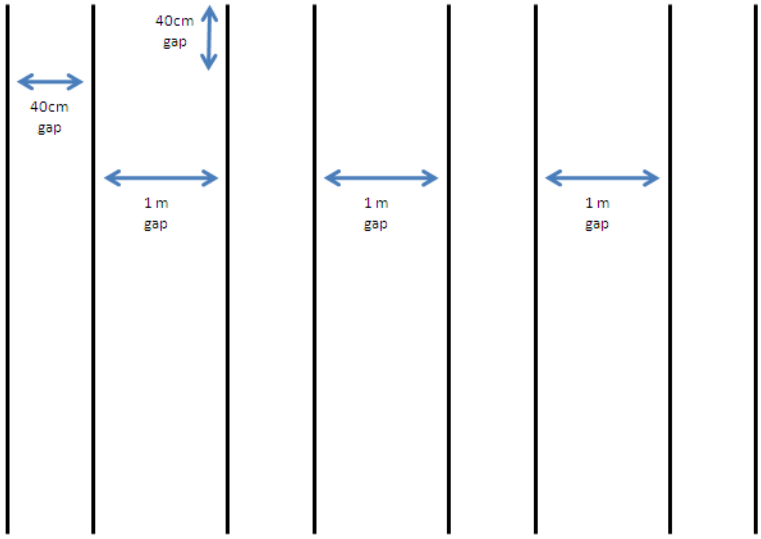


Figure 4.4 – Field trial layout – Plants transplanted in rows; 40cms apart in two rows, before a gap of 1m and a further two rows separated by 40cms

Saladin tends to 'head up' i.e. form tightly compacted heads as it develops. The heads were cut transversely at the appropriate time to enable the bolt to emerge and prevent the lettuce from rotting, see figure 4.5, any basal leaves making contact with the ground were removed to prevent disease. Leaf material was collected for screening at the molecular level as described in section 2.1.2.3.

Figure 4.5 – Saladin plant in the field - to avoid disease base leaves touching the ground are removed. The head of the plant has been cut to allow the bolt to emerge, reducing the likelihood of it rotting

4.3 Materials and Methods - Larissa Experiments

4.3.1 EMS Treatment of seeds

The Larissa population was created at Warwick HRI by Prof. Dave Pink's research group. Larissa seeds (M0), were soaked in 500 seed lots in 0.1 M phosphate buffer pH 7 at 25 °C containing 0.1 %-0.4 % EMS (Sigma, Cat. No. M0880), in a shaking incubator at 80 x g for 24 h. Seed was then rinsed in 100 mM sodium thiosulphate to inactivate the EMS. M1 and M2 seed was grown and M3 seed was collected prior to work on this thesis beginning. A mixture of residual M2 seed remaining from previous experiments and M3 seed was made available for this study, the mutant population was made up of 62 lines.

4.3.2 Screening of a Larissa EMS Population for bolting time

12 plants from each line (where possible M2 seed, but if unavailable M3 seed), was used, 12 WT plants were also grown, along with 12 plants from 3 sets of control seed that had been treated with sterile water rather than EMS. The seeds were sown and transplanted as described in section 2.1.2.1. The growing conditions were as described in section 2.1.2.3. This experiment took place in the glasshouse during the winter months, 16 h LD conditions were provided by supplementary lighting. Plants were scored for bolting as described in section 4.1.2.

Four seeds collected from each of the selected lines with interesting EB and LB phenotypes were re-scored in natural LD photoperiods over the summer in the glasshouse to ensure the robustness of the bolting phenotype. Those plants replicating the desired bolting phenotype were used in the first BC experiments.

4.3.3 Backcrossing of LB Larissa lines

To clean up the mutations causing the LB phenotype from other non-advantageous background mutations induced by the EMS treatment, the lines were BC twice to Larissa WT plants, figure 4.6 illustrates the strategy used for this project. The method used to BC the plants is described in section 2.2.16.

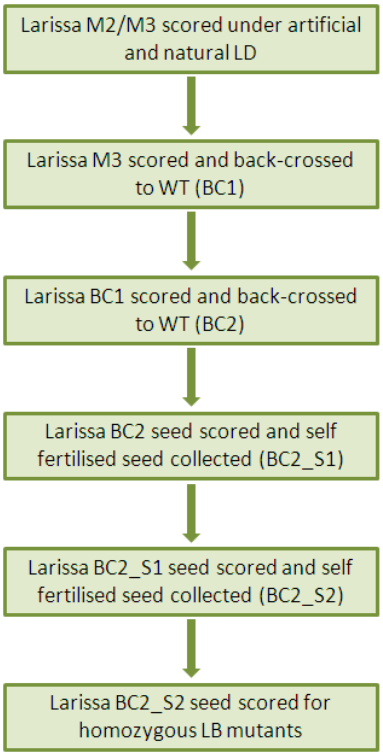


Figure 4.6 – Strategy undertaken to obtain homozygous LB Larissa lines

Plants used in the BC experiments were scored to ensure they were LB under natural LD conditions of 16 h in the glasshouse. Four plants per line were then BC to WT, an average of 20 BC events took place per plant. Seed was collected, and for the purposes of this report, this seed will be referred to as back-cross 1 (BC1) seed.

Four seeds, one representing each of the four BC1 plants was sown and scored for number of days to bolt under natural LD conditions in the glasshouse, before being back-crossed to Larissa WT for a second time. This seed was collected and will be referred to as back-cross 2 (BC2) seed.

Four seeds from one BC2 event for each of the four BC1 plants grown was sown (16 plants in total), and scored, for number of days to bolt under natural LD conditions in the glasshouse. To try to obtain homozygous mutant lines with a LB phenotype, these plants were allowed to set self seed, referred to as back-cross 2 selfed 1 (BC2_S1) seed.

Four BC2_S1 seed from each of the 16 BC2 plants were sown and scored, under artificial LD conditions in the glasshouse, for the number of days to bolt. Selfed seed was collected and will be referred to as back-cross 2 selfed 2 (BC2_S2) seed.

20 BC2_S2 homozygous mutant seeds from lines with an interesting LB phenotype were then grown and scored under natural LD conditions in the glasshouse to confirm the LB bolting phenotype. In a second experiment a further 40 BC2_S2 plants representing the interesting LB lines were scored under natural LD conditions in the glasshouse, 20 seedlings were subjected to a vernalisation treatment, as described in section 2.1.2.3, the other 20 seedlings were not treated.

One of the aims of this project was to increase the holding ability of lettuce plants in the field. All experiments to this point had taken place in the glasshouse. Therefore, to test whether these lines were still LB in the field, the LB homozygous mutant lines identified through these experiments were grown in a field trial. Two field trials took place in Wellesbourne, Warwickshire, UK, the second trial was replicated at JE Piccavers farm in Spalding, Lincolnshire, UK, 52°47', see figure 4.7. The plants in the first of the two trials were transplanted 11th May 2010 and for the second trial were transplanted six weeks later, in the week beginning 21st June 2010.

Figure 4.7 – Verification of the Larissa lines with a LB phenotype in the field. Photos show the two trial sites, the left photo at Wellesbourne and the photo on the right in Spalding.

40 WT plants and 20 plants representing six Larissa homozygous mutant lines were sown and transplanted as described in section 2.1.2.1. The plants were randomly transplanted into rows as illustrated in figure 4.4, (the plants used in the Lincolnshire trial were transplanted into four rows, each plant was 40cms apart), using a randomised complete block design balanced in both rows and columns at the main plot level (Mead *et al*., 1993). Watering was applied when required, but no leaves were removed from the plants in this trial. In this experiment it was possible to monitor the development, phenotype and biomass of the mutant lettuce plants in direct comparison with WT plants.

4.3.4 Investigating the effect of temperature on LB Larissa lines

Temperature has been shown to be a major contributory factor affecting bolting time in lettuce. To test the effect of temperature on the LB Larissa lines identified, plants were grown under four different temperatures to investigate the effects of higher temperatures on the bolting times of the LB lines identified. Seeds were sown and transplanted as described in section 2.1.2.1. Four plants from each of the LB lines and WT were grown in growth cabinets, see figure 4.8, at 18 °C, 21 °C, 23 °C and 25 °C under 16 h LD of white light (100 $\mu mols^{-1}m^{-2}$). The temperatures chosen were to cover the average UK temperature range recorded from May-August of 18 °C - 23

°C, (Met Office, UK), plus a higher temperature of 25 °C to investigate the effect that climate change, and warmer summers may have on the lines. The number of days to bolt at each temperature was recorded.

Figure 4.8 – Experimental set up for temperature experiments. Four plants from each of the mutant lines and WT were grown at four different temperatures in growth cabinets.

Appendix VII table A5, contains a complete breakdown of all the Larissa experiments completed and the time frame of each experiment in the context of the project.

4.3.5 Statistical analysis of the Larissa population

As described in section 4.3, one-way ANOVA tests were applied to the bolting data collected. The test was applied to the original screen of the population and to experiments based on the homozygous LB lines identified after BC and selfing. ANOVAs were not applied to the data obtained from generations of plants at which point the mutations were segregating, as any analysis performed would skew the data. A better way of analysing this data was to look at each line individually to assess the ratio of LB plants to plants bolting as WT. Tables corresponding to all ANOVA values quoted in this thesis can be found in Appendix VIII tables A7-A52.

Because the trait we are concerned with; LB, is a very subjective score, the ANOVA values have been used as a guide. Under some environmental conditions (LD photoperiods, increased temperatures) plants bolting just two to three days later than WT can be classified as LB. As a general rule plants grown in the glasshouse, unless otherwise stated, were classed as LB if they bolted on average >7 days or more after than WT. The other constraint that needed to be considered was the amount of space available to run the bolting trials; on occasions a significant ANOVA value will include more lines than it was possible to accommodate.

4.4 Materials and Methods - Saladin methods

4.4.1 EMS treatment of seeds

Two independent batches of seed treated with EMS made up the Saladin population. An initial 'test' batch of seed was set up to establish the appropriate EMS concentration/treatment to use, these seeds were treated as follows; seeds were treated with 0.1 % EMS phosphate buffer at 25 °C for 24 h, as described in section 4.3.1. M1 plants were then grown before work on this thesis commenced.

4.4.2 Screening of Saladin EMS populations for bolting time

4.4.2.1 'Test' Population

Seeds from the 'test' population made available for this study were from the M2 generation, which comprised of 125 lines. Four seeds from each M2 line were sown and transplanted into the haygrove tunnels as described in section 2.1.2.1. 20 WT plants were also sown, these plants were randomly placed throughout the haygrove tunnels. The plants were maintained as described in sections 4.2.1 and were scored for number of days to bolt, as described in section 4.1.2. The M2 generation of this

population was screened twice in the haygrove tunnels over two consecutive summers. M2 seed from interesting lines identified from the first haygrove screen were also grown in the glasshouse, under natural LD photoperiods to verify the LB phenotype and to BC to WT. Four plants from each line were sown and transplanted as described in section 2.1.2.3, plants displaying a LB phenotype were BC to WT as described in section 2.2.16. Seed (M2_BC1), from these plants was collected and where seed was available, six plants from each BC1 event was sown and BC to WT for a second time. Seed (M2_BC2), from these plants was collected and was sent to Rijk Zwaan®, Fijnaart, Holland for selfing over the winter period under artificial lamps, set to create a LD photoperiod of 16 h. Eight seeds from two BC2 events from two plants were sent and selfed. The selfed seed obtained back from Rijk Zwaan® (M2_BC2_S1), was sown in the glasshouse under natural LD conditions and was scored for number of days to bolt, six seed from two independent BC2 events from two BC2 plants (24 plants) were scored along with 30 WT plants.

4.4.2.2 'Main' Population

The second 'main' Saladin population was created as per the test population, except the seed was treated for 18 h rather than 24 h. The M1 population was made up of 2612 individual lines which were sown and transplanted as described in section 2.1.2.1, due to the large number of lines in the study no randomisation of the plants took place. The plants were scored for the number of days to bolt, 20 WT Saladin plants were also grown and scored; these plants were transplanted in a randomised manner throughout the haygrove tunnels.

M2 seed was collected from 1012 of the M1 plants from the 'main' population. Four M2 plants were grown from seed collected from each M1 plant, as described in

section 4.2.1. These plants were scored, along with 20 WT Saladin plants, for the number of days to bolt. M3 seed was collected from these plants.

Four M3 plants from the lines identified as LB in the M2 population were grown and scored for number of days to bolt in the glasshouse. Lines displaying a LB phenotype were BC to Saladin WT, as described in section 2.2.16. To maximise the chances of obtaining homozygous mutant lines, self fertilisations from the BC1 seed collected were set up, the Saladin self fertilisations discussed in this thesis were all performed by Rijk Zwaan®, Fijnaart, Holland, over winter under artificial lamps. The BC1 seed collected from two independent events from two individual plants was self fertilised. The seed collected, BC1_S1, was then screened for homozygous LB mutant lines. Six plants representing the self seed collected from the two independent BC1 events from two individual BC2 plants (24 plants) from each individual LB line were grown in the glasshouse under a natural LD photoperiod and were scored for days to bolt.

The strategy utilised in this project to produce homozygous Saladin mutant lines with a LB phenotype is shown in the flow diagram in figure 4.9.

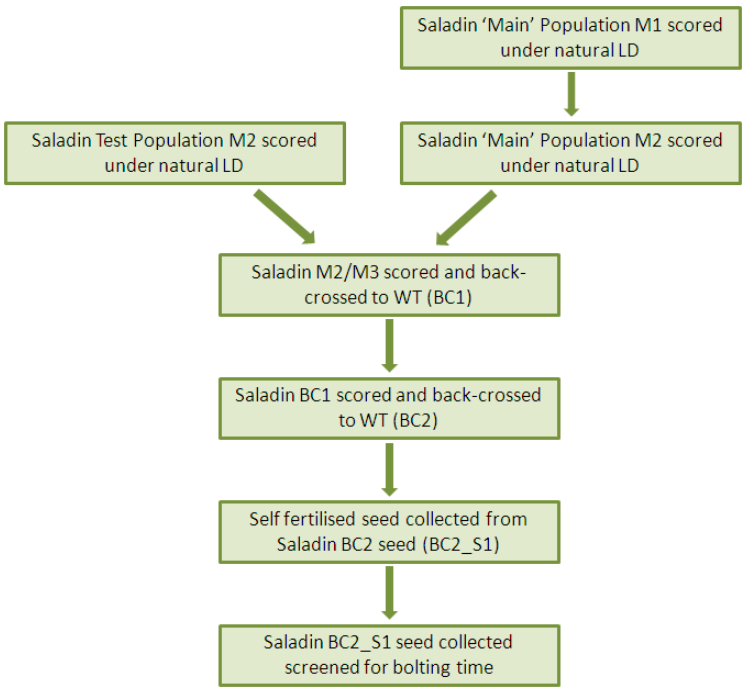


Figure 4.9 – Strategy undertaken to obtain homozygous LB Saladin lines

Appendix VII table A6, contains a complete breakdown of all the Saladin experiments completed and the time frame of each experiment in the context of the project.

4.4.3 Statistical analysis of the Saladin populations

ANOVA tests and Q-Q plots were calculated for the Saladin populations as described in section 2.3 and in section 4.3.5. Again these values were used as a guide to decide which plants were scored as LB. Unless otherwise stated plants bolting two weeks or later than WT (covering roughly 20 % of the plants in each trial) were classed as LB. Plants which bolted ten days later than WT in the glasshouse were scored as LB. On one occasion the plants were grown in the mid-summer in the glasshouse and all plants bolted considerably earlier than expected, ANOVA scores were utilised on this occasion.

4.5 Results and Discussion – Larissa Population

4.5.1 Photoperiodic/Vernalisation Responses of L. sativa cv. Larissa and cv. Saladin

Two sets of 12 Larissa WT, and 12 Saladin WT, seeds were grown in the glasshouse over winter. One set was placed under natural SD conditions, and one set was placed in an artificial, 16h LD photoperiod, with supplementary lighting provided when light levels dropped below a set threshold to extend the photoperiod to 16h. Plants of both varieties grew quicker and were healthier under LD conditions see figure 4.10. Larissa plants grown in a LD photoperiod bolted after an average of 77.6 +/-1.15 days; those grown in a SD photoperiod began bolting after 140 days, which may have been caused by the increase in daylength as winter turned into spring. The Saladin plants did not bolt under either SD or LD conditions with the supplementary lighting, suggesting that Saladin may have a light quality requirement to enable the

transition to flowering. External radiation levels were recorded throughout these experiments and an average level of just 2.3 W/m^2 was recorded throughout.

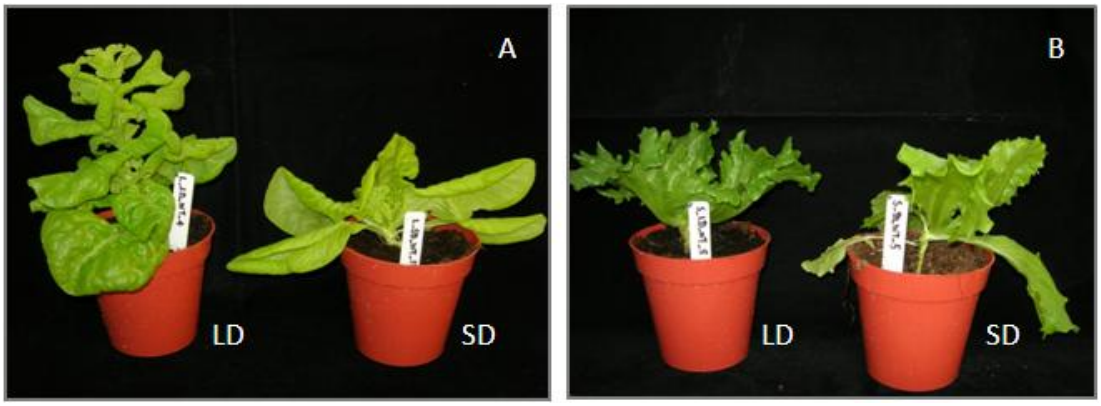


Figure 4.10 – Examples of lettuce plants grown under different photoperiodic conditions and their resultant growth responses. Photos taken 92 days after sowing.
(A) - Larissa plant development; plants grown in LD conditions develop more quickly than plants grown in SD conditions. The plant grown in LD conditions has bolted.
(B) - Saladin plant development; plants grown in LD conditions have developed more quickly than plants grown in SD conditions. The plant grown in LD conditions, although not yet bolted is more developed and looks healthier.

Larissa and Saladin plants were subjected to a vernalisation treatment of four weeks at 4 °C and bolting times in the above LD and SD conditions was recorded. No significant differences in days to bolt were observed between vernalised and non-vernalised plants. This suggests that neither cultivar has a vernalisation response or else it is one that cannot be satisfied by four weeks cold treatment.

4.5.2 Screening of L. sativa cv. Larissa EMS mutagenised population for late-bolting lines

4.5.2.1 M2 / M3 Screen

M2 and M3 seed from 62 individual Larissa lines making up the EMS mutagenised population were made available by Dr. Paul Hand, Warwick HRI. 12 plants from each line were grown in a glasshouse under an artificial 16h LD photoperiod.

A good indication that the EMS treatment had been successful was the various phenotypes observed in the population of lettuce plants grown. Figure 4.11

illustrates some of these phenotypes, including changes in leaf morphology. Bolting was quantified when flowering parts, e.g. floral buds, were observed see figure 4.2 in section 4.1.2.

Figure 4.11 – Phenotypic variation observed in EMS mutagenised M2 and M3 Larissa plants

A range of bolting times for the Larissa population was observed, figure 4.12. The earliest plants bolted after just 50 days, with the latest plants bolting after 121 days. All WT plants bolted between 65 and 82 days.

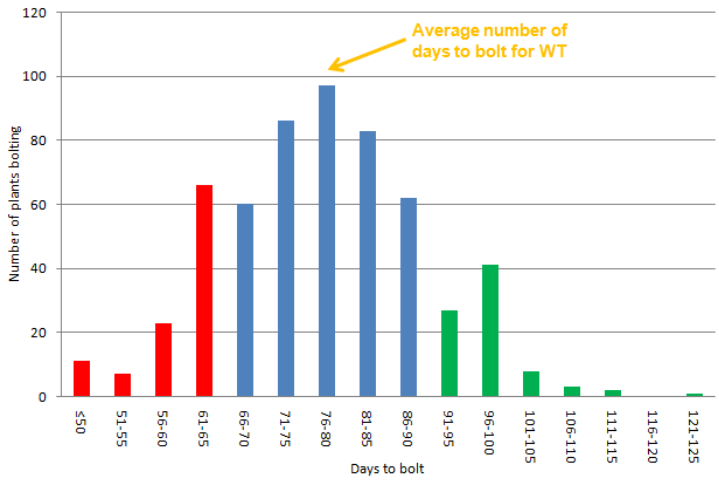


Figure 4.12 – Range of bolting times in Larissa EMS population. Larissa WT bolted between after an average of 77.6 +/-1.15 days, highlighted in orange. Plants classified as EB and LB are highlighted in red and green respectively.

Figure 4.13 shows the variation in the number of days to bolt of three individual plants. The EB plant bolted after just 50 days and had flowered and was setting seed as well as showing signs of leaf senescence when the photo was taken 99 days after sowing. In comparison the LB plant did not bolt until day 101.

Figure 4.13 – Range of bolting times identified from Larissa M2/M3 EMS mutagenised population. The EB plant bolted after 50 days, the plant representing an average number of days to bolt (Ave), bolted after 79 days and the LB plant went on to bolt after 101 days. Photo taken 99 days after sowing.

Figure 4.14 shows the average bolting time of the 12 plants for each Larissa line. A one-way ANOVA was calculated to assess the significance of the difference in number of days to bolt per line for the Larissa population. L.s.d. values (P=0.005) were calculated and EB and LB lines were identified with a significant difference in average bolting time to the non-treated (orange bar) and water-treated control lines (purple bars). Five lines bolted significantly earlier than the controls, (red bars), and seven lines bolted significantly later, (green bars), these were taken for further analysis (p=<0.001; d.f.=524; l.s.d.=12.69).

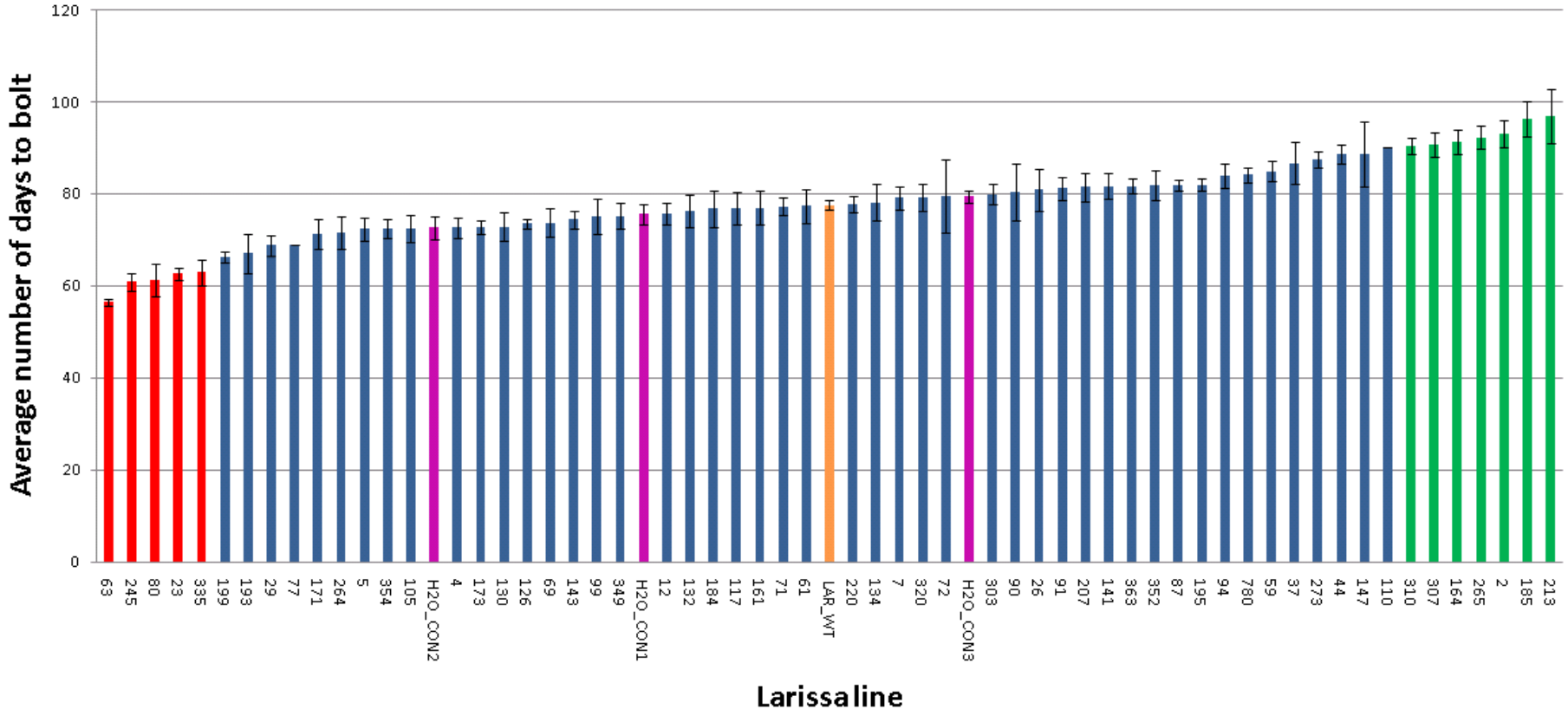


Figure 4.14 - Average bolting time in days for individual EMS mutagenised Larissa lines. The orange bar shows the Larissa WT control plants. Red and green bars refer to the EB and LB plants. Purple bars indicate water treated control lines.
Error bars for ALL graphs in chapter 4 refer to the Standard Error of the Mean.

A Q-Q probability plot was constructed to look for variation occurring within the Larissa population. Figure 4.15, shows the output from the Q-Q analysis, each cross represents the average number of days to bolt for each individual line, if the distribution of the population is similar then each cross will lie along the green line. Each of the lines plotted occur within the red lines which illustrate the 90 % confidence limit that the data fits a Gaussian distribution. Because the data fits a Gaussian distribution we can calculate that there is only a 10 % chance that the seven LB lines would have bolted after an average of 90 days or more and only a 7 % chance that the five EB lines bolt after an average of 62.7 days or earlier. This confirms that the lines chosen are significantly different from the rest of the population screened and provides confidence that the phenotypes observed are less likely to be caused by natural variation, indicating something else is contributing to the bolting phenotypes observed.

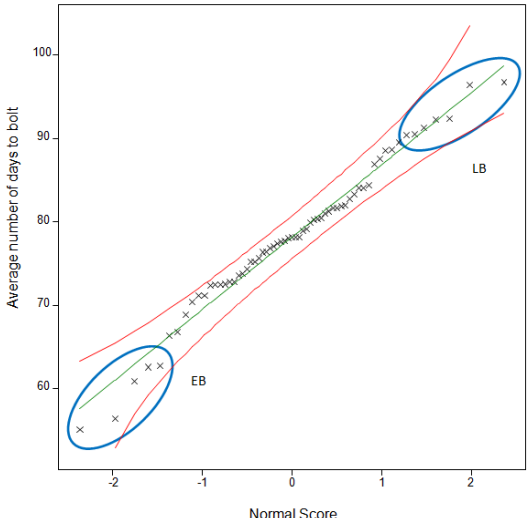


Figure 4.15 - Normal Q-Q probability plot for trt (+ 90 % Confidence Limits) for Larissa EMS population. The skewed red lines represents the 90 % confidence level that the data fits a Gaussian distribution, the lines highlighted in blue are the EB and LB lines, these phenotypes are less likely to be caused by natural variation, indicating something else is causing the bolting phenotype observed.

Therefore the lines chosen to follow up as LB bolted after an average of 90 days or more. The lines which bolted after 65 or fewer days were designated EB. A difference of over two weeks exists in days to bolt when comparing the untreated

control plants and the earliest and latest bolting lines. The earliest line to bolt, line 63, bolted on average after only 56.4 +/-0.73 days, compared with WT which bolts on average after 77.6 +/-1.15 days. The latest bolting line, line 213, bolted on average after 96.83 +/-5.87 days.

Interestingly some of the plants that bolted earlier than WT did not go on to produce flowers that opened earlier than WT. It appears that the number of days that a plant takes to bolt is not directly related to the number of days it will take to open its first flower.

4.5.2.2 Confirmation of the bolting phenotypes under natural LD light conditions

Seed was collected from those plants displaying the EB or LB phenotype, no seed was produced by plants representing one EB (line 63), and one LB line (line 213), possibly due to background EMS mutations affecting seed production, these lines were lost from the study at this point, see table 4.1 for details. All seed collected is either M3 or M4 seed.

EARLY BOLTING LINES			LATE BOLTING LINES		
Line	**Mean Days to Bolt**	**Seed Availability**	**Line**	**Mean Days to Bolt**	**Seed Availability**
M3_63	56.4	No seed produced	M2/M3_185	100.7	Seed collected
M2_245	60.9	Seed collected	M2_213	96.8	No seed produced
M2_80	61.2	Seed collected	M2_2	93	Seed collected
M2_23	62.5	2 seeds available	M2_265	92.3	Seed collected
M2_335	62.9	2 seeds available	M3_164	91.3	Seed collected
WT	77.6	N/A	M2/M3_307	90.7	Seed collected
			M2_310	90.5	Seed collected
			WT	77.6	N/A

Table 4.1 – Larissa EB/LB identified from screening M2/M3 plants. M3/M4 seed was collected where available for future experiments

To confirm these lines were displaying interesting phenotypes, four plants from each line were grown in the glasshouse under a natural LD light photoperiod. Where possible seed from the same stock used in the initial screen was used in the re-screen, however some of this seed had been used up and M3 or M4 seed was used. Interestingly plants from all lines bolted much earlier when grown over the summer months under natural light conditions than when grown over winter months with supplementary lighting. The average external radiation levels for plants grown over the winter period was much lower than for those grown over the summer months; an average external radiation level of 2.6W/m^2 was recorded throughout December compared to an average value of 23.2W/m^2 throughout June. The EB lines appeared to lose the EB phenotype when grown over the summer months, however, five of the six LB lines; 2, 164, 265, 307 and 310 still bolted significantly later than WT

(p=<0.001; d.f.=17; l.s.d.=7.641), on average ≥1 week later than WT see figure 4.16, ANOVA confirms that the five lines mentioned are bolting significantly later than WT and that line 185 is no longer displaying a LB phenotype; on average this line bolts only 4.6 days later than WT. As the EB lines and the LB line 185 did not show a reproducible phenotype under a natural LD photoperiod, no further analysis was done on these lines.

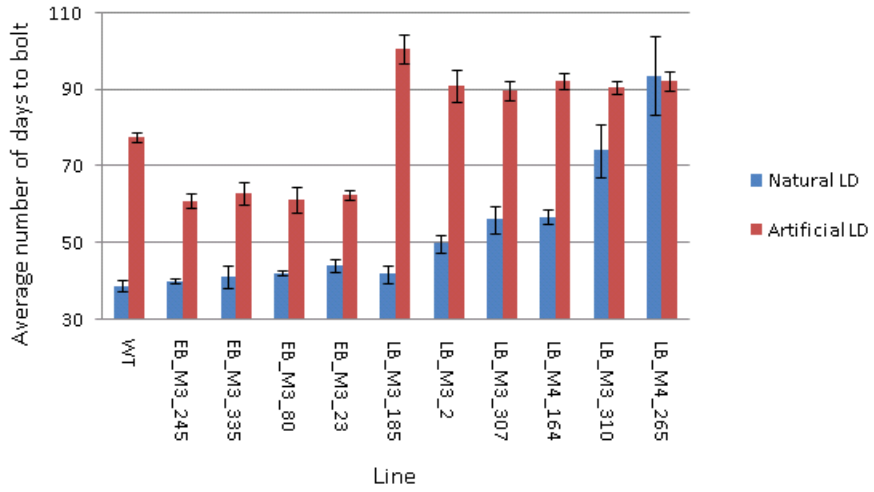


Figure 4.16 – Comparison of bolting times in Larissa lines grown under artificial and natural LD light conditions. EB lines lose the phenotype under natural light, whereas five of the LB lines maintain the phenotype – lines 2, 164, 307, 310 and 265.

Of the five LB lines that were late in both experiments lines 265 and 310 displayed abnormal phenotypes caused by the EMS treatment which were very likely to have an effect on the bolting phenotype. Line 265 plants were very weak/flimsy plants whereas plants from line 310 were very small, see figure 4.17. This project is only interested in mutations that specifically affect flowering/bolting and not ones that have pleiotropic effects, these lines were discounted from the study at this stage.

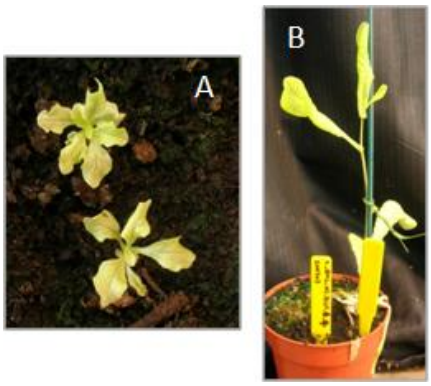


Figure 4.17 – Phenotypes of Larissa EMS mutagenised lines 310 (A), and 265 (B). Both lines have undesirable phenotypes caused by the EMS treatment. The poor development of these lines may be the cause of late-bolting.

Lines 2, 164 and 307 all displayed a normal phenotype, however because the M1 seed from these lines was not available for this study it was difficult to ascertain whether the mutation(s) causing the LB phenotype are dominant or recessive. Figure 4.18 shows the data collected from the three lines (2, 164 and 307), in the original screen, individual plants from the LB line are highlighted in green, WT plants are in orange. Not all of the 12 plants grown for each mutant line survived to be scored for bolting.

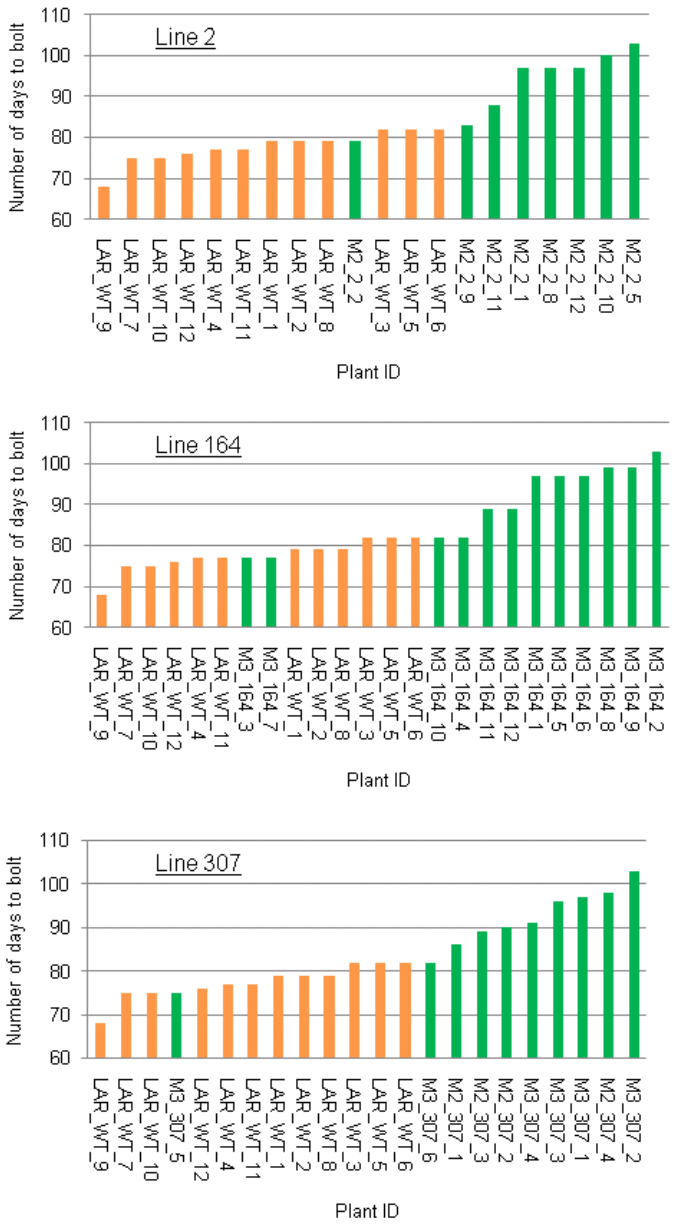


Figure 4.18 – Number of days to bolt for each plant making up the LB Larissa lines 2, 164 and 307. Orange bars represent WT plants and green bars indicate mutated Larissa lines.

To try to explain the nature of the mutation causing the LB phenotypes the bolting data from each line has been analysed further. Eight of the 12 M2 plants sown from line 2 survived to the bolting stage, six of which bolted significantly later than WT (seven days or more). As these plants are from the M2 generation, and therefore the plants should be segregating for the mutation at this stage, it would be expected that if the phenotype was caused by a dominant mutation then a 3:1 ratio of LB plants to plants bolting as WT would be observed. Six of the eight plants were LB, a 3:1 ratio.

Furthermore three of the four M3 plants, collected from a single LB M2 individual, scored in the experiment verifying the bolting phenotype under natural LD conditions were also later than WT. All four M3 plants should be late if the M2 parent plant was homozygous for the mutation, however if it was heterozygous and the mutation was dominant then a 3:1 LB to WT ratio would again be expected. So whilst a larger number of plants in this experiment would have been more conclusive, it is likely that a dominant allele is causing the LB phenotype.

There was no M1 or M2 seed available for line 164 as it had all been used in previous experiments, 12 M3 seeds collected from a single M2 plant produced eight plants with a LB phenotype and four plants which bolted as WT, see figure 4.18. A 3:1 ratio of LB plants to plants bolting as WT would be expected if a dominant allele was causing the LB phenotype, a 2:1 ratio is seen in line 164. We can infer that the M2 plant from which the M3 seed was collected was heterozygous because a homozygous M2 plant would have resulted in all the M3 progeny having a LB phenotype or a WT bolting time. If a recessive allele was causing the LB phenotype a 3:1 segregation of plants bolting as WT to LB individuals would be seen in the M3 generation. As the M3 population has a 2:1 LB individuals to plants bolting as WT, which is closer to 3:1 LB to WT than 1:3 LB to WT, it would appear that the mutation causing the LB phenotype is a dominant one. However due to the relatively small number of plants scored and the possibility that potentially more than one mutation is affecting bolting time, this cannot be certain. For line 307 of the four M2 plants that were planted in the first screen, all bolted later than WT, suggesting that a dominant mutation may be causing the LB phenotype. Of the eight M3 seeds sown in this screen, only six plants germinated; two bolted as WT, with the other four bolting late, see figure 4.18. The four M3 plants, derived from a LB M2 plant, rescored over

the summer months under a natural LD photoperiod all displayed a LB phenotype. As with lines 2 and 164 we can infer that the mutation(s) causing the LB phenotype in line 307 is dominant. A schematic explaining the possible genetics occurring in the LB Larissa lines 2, 164 and 307, is shown in figure 4.19.

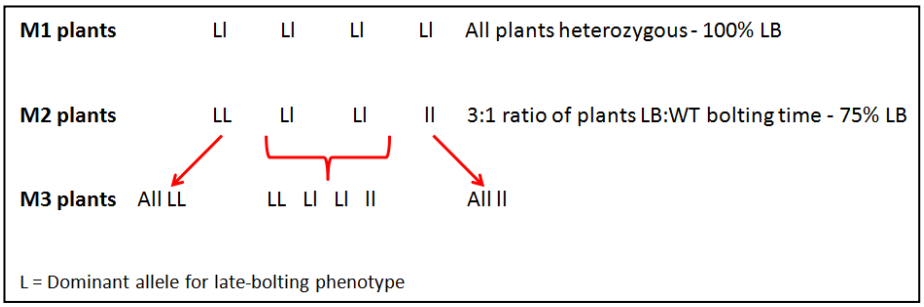


Figure 4.19 – Possible genetics occurring in the late-bolting Larissa EMS mutagenised lines

4.5.2.3 Removal of undesired alleles generated by EMS treatment

In an attempt to clean up the mutation(s) causing the LB phenotype from other EMS induced background mutations, in the three lines of interest, two rounds of BC to Larissa WT plants were performed. BC was carried out in the glasshouse, as described in section 2.2.16. Four seeds, collected from individual LB M2/M3 plants were grown, and these M3/M4 plants were BC to Larissa WT to produce, back-cross 1 seed (BC1). Four BC1 seeds, one from each of the four BC plants per line were grown and BC to Larissa WT for a second time to produce back-cross 2 seed (BC2). Self fertilisations were set up by sowing four seeds from four independent BC2 events, i.e. 16 seed derived from each BC1 event, to maximise the chances of obtaining homozygous LB mutant lines. Self seed (BC2_S1) was collected from these plants and four BC2_S1 plants were grown from each selfed plant, resulting in 64 plants being screened for bolting time for each of the three LB lines. The BC and self-fertilisation strategy is illustrated in figure 4.20.

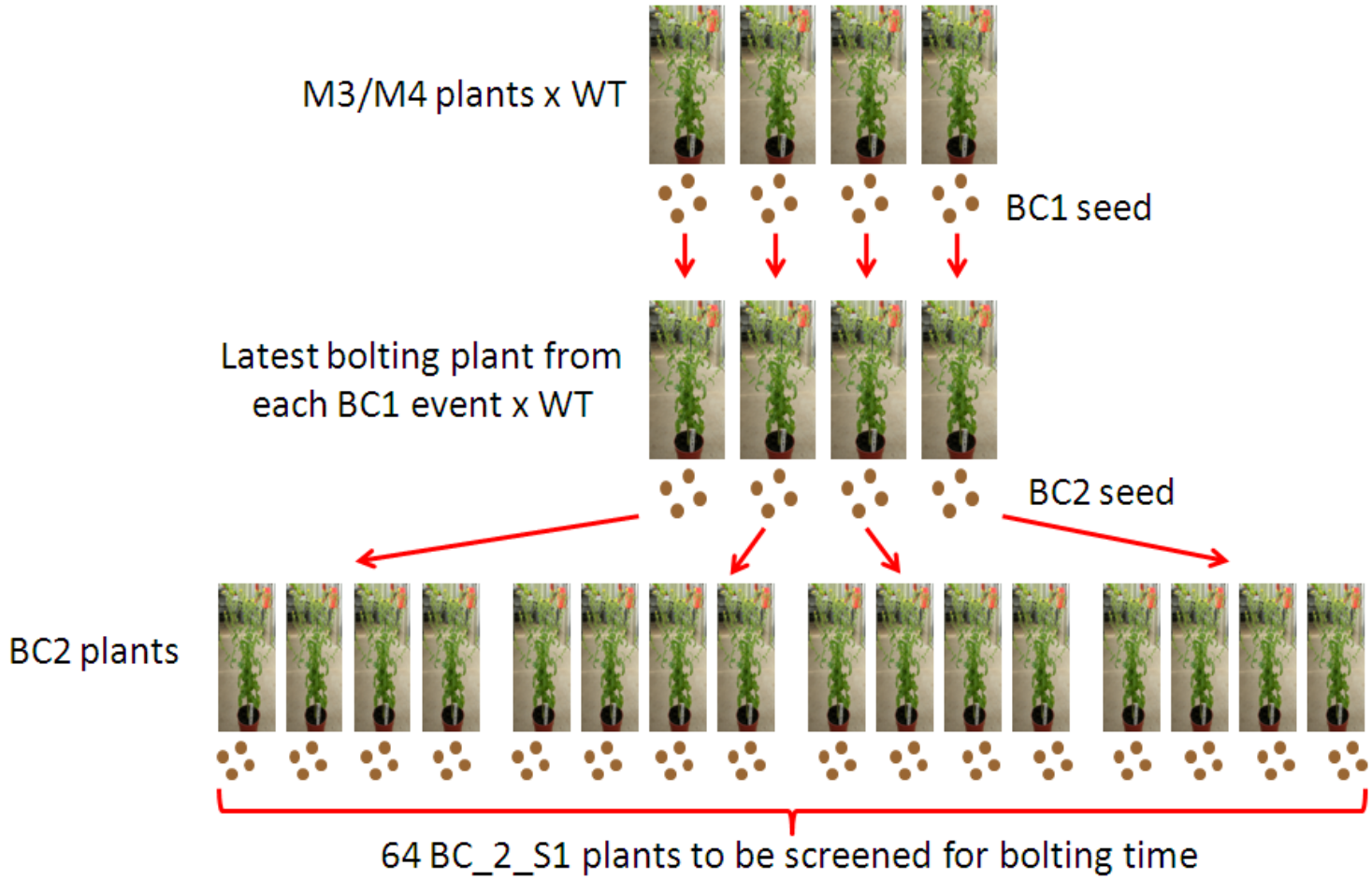


Figure 4.20 – Explanation of the BC and selfing strategy undertaken. A second round of BC takes place using BC1 plants, then four seeds from four independent events are grown and self-fertilised. Four seed from each self-fertilised plant were scored for bolting time – resulting in 64 plants being scored in total.

The plants used in each of the BC experiments were scored to attempt to ascertain the way in which the mutation(s) causing the LB were segregating throughout the population. The four original plants used for BC1 experiments to WT were LB as described in section 4.5.2.2 and displayed in figure 4.16. The four BC1 plants derived from four independent BC1 events were also scored for days to bolt, see figure 4.21. All BC1 plants from line 2 bolt later (two to ten days), than the average days to bolt recorded for WT. This is consistent with it being a dominant mutation that is causing the bolting phenotype because a dominant homozygous mutation in the M3 plants crossed with WT would result in four heterozygous plants in the BC1 generation all with a LB phenotype; see figure 4.21. Alternatively it is also possible that the LB phenotype observed in the BC1 plants could mean that the BC has not been successful, if self pollen from the mutant line is not washed from the stigmas then a self-fertilisation event will occur resulting in selfed seed. It was to minimise this risk that plants from four independent BC1 events were used for the second BC. Lines 164 and 307 are slightly different, most of the individual BC1 events resulted in two plants with LB phenotypes and two with WT phenotypes, this again suggests that the mutation causing the LB phenotype is dominant but is heterozygous in the M3/M4 plants. A BC to WT in this case would result in 50 % of plants displaying a LB phenotype and 50 % with a WT phenotype, see figure 4.21.

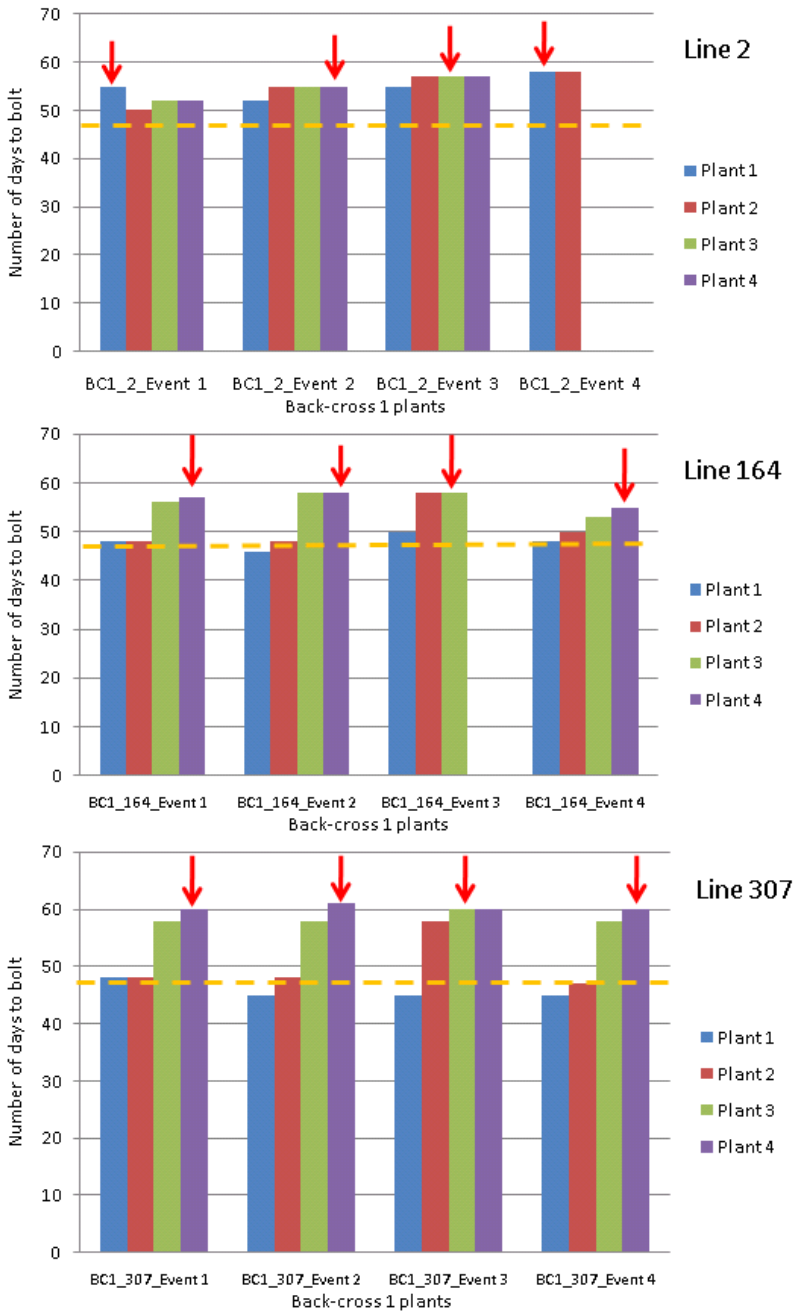


Figure 4.21 – Number of days to bolt of BC1 plants from lines 2, 164 and 307. Four plants from four independent BC1 events were scored for number of days to bolt. The orange dotted line represents the average number of days to bolt for WT. The red arrows indicate the plants taken to use for the second round of BC.

From the four plants derived from each BC1 event, the latest bolting plant (indicated with red arrows in figure 4.21) was used for the second round of BC to WT and the production of BC2 seed. As the BC1 plants would be heterozygous for the mutation, the BC2 seed collected from all three lines should be segregating for the dominant LB mutation. This should result in 50 % of the BC2 plants being LB and 50 %

bolting as WT, see figure 4.24 for details of the segregation of the LB allele. Four seeds from each of the four BC2 events per line were grown under a natural LD photoperiod, to score bolting, and to produce selfed seed. Figure 4.22 displays the days to bolt recorded for each of the lines. As predicted most BC2 events have resulted in approximately 50 % of plants from each BC event bolting later than WT and 50 % bolting as WT. Self seed (BC2_S1) was collected from each of the BC2 plants.

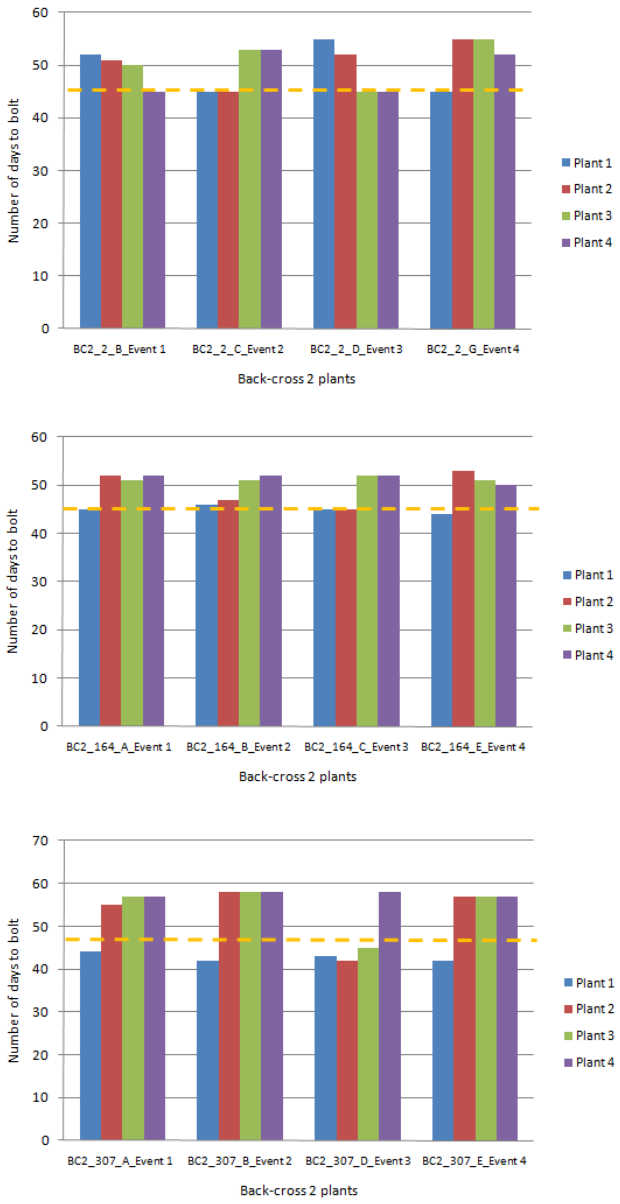


Figure 4.22 - Number of days to bolt for BC2 plants from lines 2, 164 and 307. Four plants from four independent BC2 events have been scored for number of days to bolt. The orange dotted line represents the average number of days to bolt for WT.

4.5.2.4 Screening the back-crossed and selfed Larissa plants for homozygous mutant lines

Four BC2_S1 plants generated from four individual selfed plants, generated from four independent BC2 events were scored under an artificial LD photoperiod to identify LB individuals. In total each of the three original LB lines (2, 164 and 307) were represented by 64 plants. Figure 4.20 illustrates the origin of the seed sown.

Eight Larissa WT plants were grown as controls, and they bolted after an average 67.4 +/-1.5 days. The 64 plants screened per LB line can be broken down into 16 groups of four, all plants within a group are selfed plants originating from a single BC2 event.

At this stage the mutation causing the LB phenotype within each line is still segregating, therefore no statistical analysis has been performed on the data, it is far more relevant at this point to analyse each line individually and look at the distribution of the LB individuals. Plants were described as LB if they bolted seven days or more after the last WT plant, this cut off point is highlighted with a green dotted line in figure 4.23.

Figure 4.23 – Number of days to bolt for each of the BC2_S1 plants generated from the LB Larissa lines 2, 164 and 307. The orange dotted line represents an average number of days to bolt for WT, the green line indicated the cut off point for LB.

The LB phenotype was not apparent in around half of the 16 groups of four screened. This was not unexpected as the mutations contributing to the LB phenotype in the original M2/3 lines could have been lost during the BC process. This was unlikely to be the case for all groups however, as they were derived from four different BC2 events. All LB BC1 plants will be heterozygous for the mutation, therefore the BC2 plants will be segregating, half will be LB heterozygous for the mutation, see figure 4.24. As all the LB mutations appear to be dominant selfing of these plants will result in a 1:2:5 ratio LL:Ll:ll or a 3:5 ratio of LB:WT phenotype where 37.5 % of the 64 selfed plants from each line would be LB compared to WT. A number of BC2_S1 lines showing a LB phenotype were obtained, nine of the 16 groups of four BC2_S1 seeds sown from line 2 produced at least one plant bolting significantly later than WT, eight of the 16 groups of four BC2_S1 seeds from line 164 produced at least one plant bolting significantly later than WT, and seven of the 16 groups from line 307 also contained at least one LB plants, the interesting groups are highlighted in green in table 4.2.

Line 2	Number of BC2_S1 plants with LB phenotype (out of 4)	Line 164	Number of BC2_S1 plants with LB phenotype (out of 4)	Line 307	Number of BC2_S1 plants with LB phenotype (out of 4)
LAR_2_BC2_S1_B1	1	**LAR_164_BC2_S1_A1**	0	**LAR_307_BC2_S1_A1**	0
LAR_2_BC2_S1_B2	0	**LAR_164_BC2_S1_A2**	1	**LAR_307_BC2_S1_A2**	0
LAR_2_BC2_S1_B3	1	**LAR_164_BC2_S1_A3**	1	**LAR_307_BC2_S1_A3**	0
LAR_2_BC2_S1_B4	0	**LAR_164_BC2_S1_A4**	1	**LAR_307_BC2_S1_A4**	1
LAR_2_BC2_S1_C1	2	**LAR_164_BC2_S1_B1**	1	**LAR_307_BC2_S1_B1**	0
LAR_2_BC2_S1_C2	0	**LAR_164_BC2_S1_B2**	0	**LAR_307_BC2_S1_B2**	0
LAR_2_BC2_S1_C3	0	**LAR_164_BC2_S1_B3**	2	**LAR_307_BC2_S1_B3**	1
LAR_2_BC2_S1_C4	0	**LAR_164_BC2_S1_B4**	0	**LAR_307_BC2_S1_B4**	2
LAR_2_BC2_S1_D1	2	**LAR_164_BC2_S1_C1**	0	**LAR_307_BC2_S1_D1**	0
LAR_2_BC2_S1_D2	1	**LAR_164_BC2_S1_C2**	0	**LAR_307_BC2_S1_D2**	0
LAR_2_BC2_S1_D3	1	**LAR_164_BC2_S1_C3**	0	**LAR_307_BC2_S1_D3**	0
LAR_2_BC2_S1_D4	0	**LAR_164_BC2_S1_C4**	0	**LAR_307_BC2_S1_D4**	2
LAR_2_BC2_S1_G1	1	**LAR_164_BC2_S1_E1**	4	**LAR_307_BC2_S1_E1**	2
LAR_2_BC2_S1_G2	2	**LAR_164_BC2_S1_E2**	3	**LAR_307_BC2_S1_E2**	3
LAR_2_BC2_S1_G3	0	**LAR_164_BC2_S1_E3**	2	**LAR_307_BC2_S1_E3**	0
LAR_2_BC2_S1_G4	1	**LAR_164_BC2_S1_E4**	0	**LAR_307_BC2_S1_E4**	2

Table 4.2 - Number of plants displaying a LB phenotype from three LB Larissa lines in the BC2_S1 generation

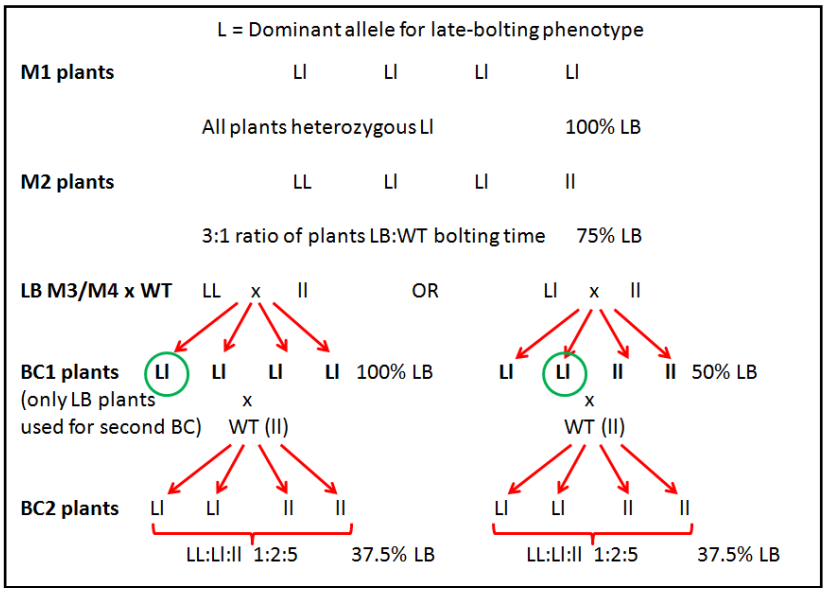


Figure 4.24 – Expected genetics of LB Larissa lines with back-crossing strategy undertaken

Seed (BC2_S2), was collected from the all of the BC2_S1 plants displaying a LB phenotype. To identify homozygous mutant lines, 20 BC2_S2 seeds of each were grown and scored for bolting time (where BC2_S1 lines contained two or more LB individuals, seed from only one of the plants was screened, due to space restrictions). From nine of the BC2_S1 LB plants representing line 2, three LB individuals produced progeny that were 80-95 % LB; bolting seven or more days later than the average number of days to bolt recorded for the WT plants screened. These lines appear to be homozygous for the LB phenotype and are highlighted in green in figure 4.25. The remaining BC2_S1 lines produced only 0-25 % LB individuals, and so are not LB.

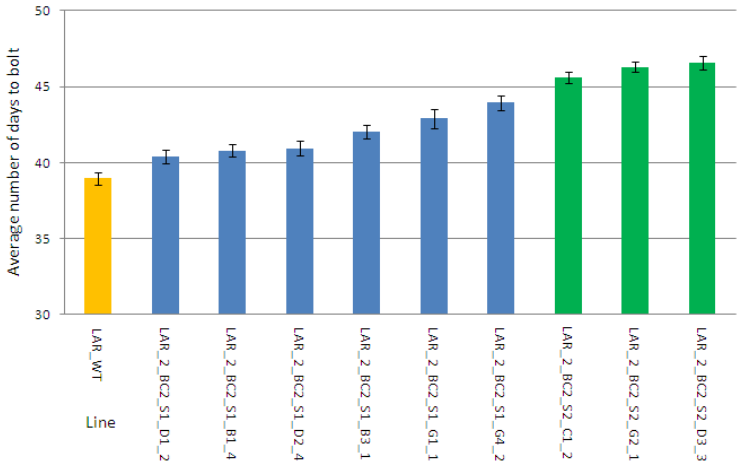


Figure 4.25 – Average number of days to bolt of 20 line 2 BC2_S2 plants. WT is highlighted in orange, lines with 80-95 % progeny bolting on average seven or more days later than WT are highlighted in green

Of the eight LB BC2_S1 plants from line 164, three produced 75-100 % LB progeny. Interestingly these three lines were all derived from the same BC1 event, the lines are highlighted in green in figure 4.26.

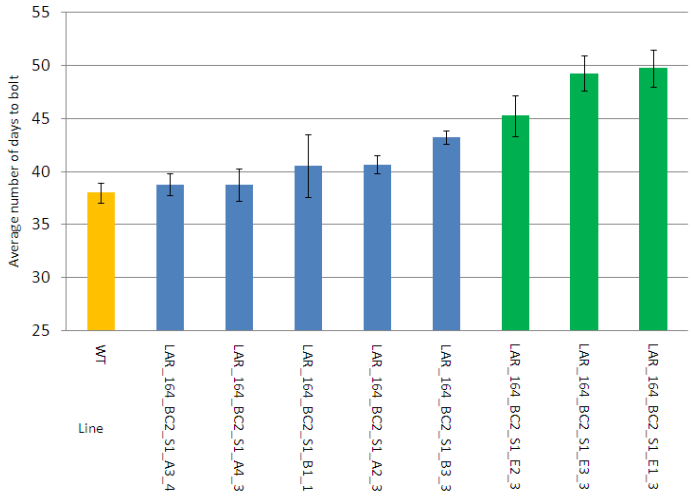


Figure 4.26 – Average number of days to bolt of 20 line 164 BC2_S2 plants. WT is highlighted in orange, lines with 75-100 % progeny bolting on average seven or more days later than WT are highlighted in green

Seven LB BC2_S1 plants from line 307 were obtained, however only one produced progeny that all had a LB phenotype (100 %), see figure 4.27, the LB line is highlighted in green. It appears that line 307 may contain more than one mutation affecting bolting time because a number of BC2_S1 and some of the BC2_S2 plants

produced plants which bolted significantly earlier than WT, most of these individuals also had an undesirable phenotype. The 307 lines were screened at a different time to lines 2 and 164, hence the difference in the average number of days to bolt obtained for WT.

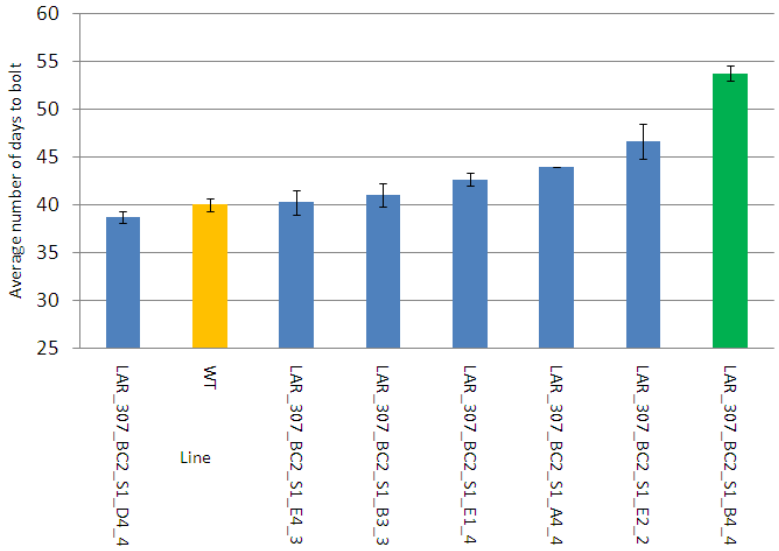


Figure 4.27 – Average number of days to bolt of 20 line 307 BC2_S2 plants. WT is highlighted in orange, all progeny from line 307_BC2_S1_B4_4 bolted on average seven or more days later than the average number of days that WT bolted after – this line is highlighted in green

4.5.2.5 Assessment of the vernalisation requirement of lines with a LB phenotype

20 seeds from the putative five homozygous LB lines from line 2, 164 and 307 (shown in green in figures 4.25 – 4.27, only one of the 164 lines verified as LB was used, as all three lines came from the same BC1 event) were resown in the glasshouse with and without a vernalisation treatment, under a natural LD photoperiod in mid-summer. The three line 2 plants were confirmed to have a significant LB phenotype under non-vernalised conditions (p=<0.001; d.f.=55; l.s.d.=1.04); plants were 5-12 days later than WT, see figures 4.28 and 4.31. A vernalisation treatment had no effect on plants from lines 2_D3_BC2_S2_3 and 2_G2_BC2_S2_1 (p=<0.001; d.f.=56; l.s.d.=1.65). There is a however a potential

effect of the cold treatment on the plants from line 2_C1_BC2_S2_2, four of the plants bolted before or at the same time as the latest WT plant, see figure 4.28.

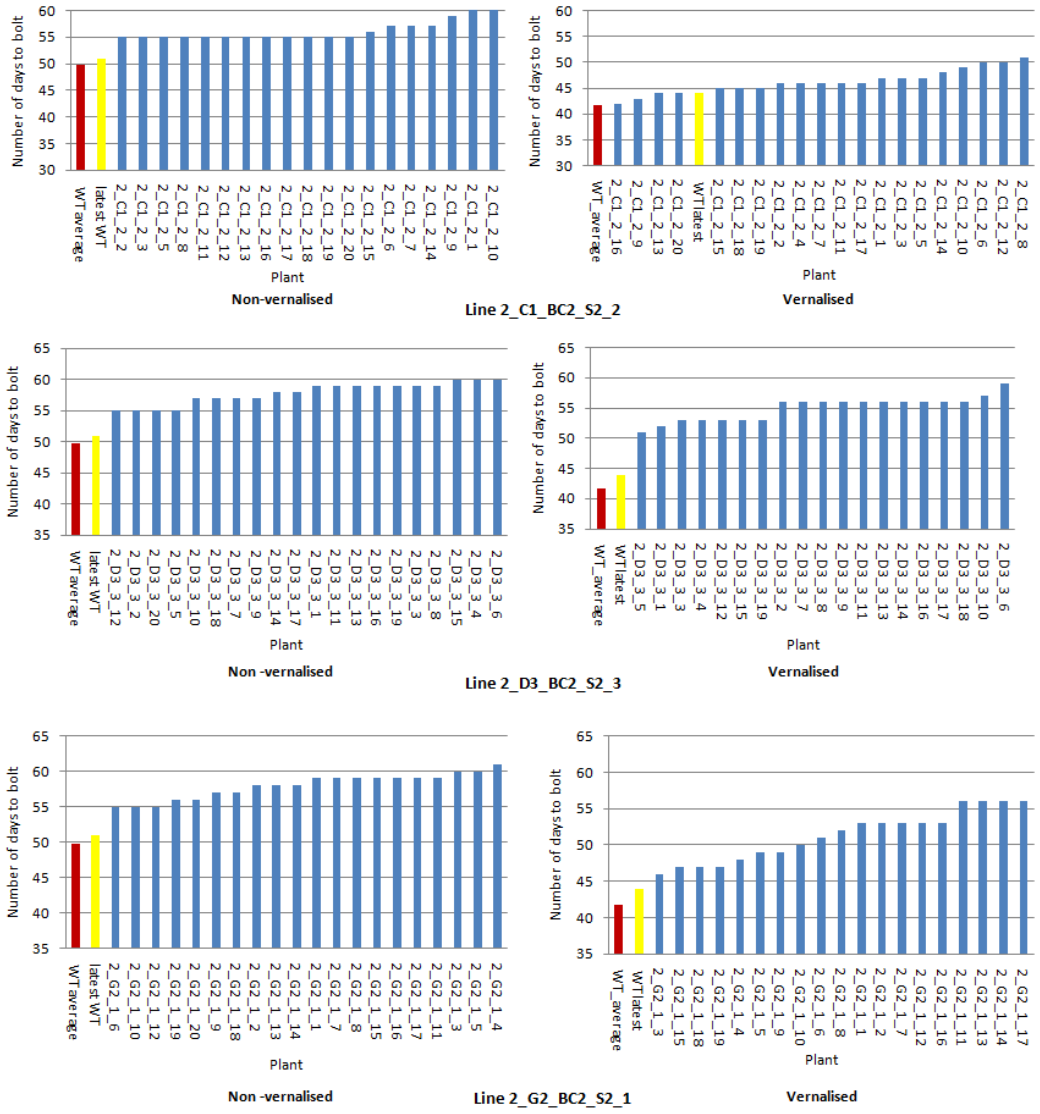


Figure 4.28 – Larissa line 2 BC2_S2 LB lines scored for days to bolt with and without a vernalisation treatment. The red bar indicates the average number of days for WT plants to bolt, the yellow bar indicates the number of days that the latest WT plant took to bolt. All line 2 plants are later than WT without a vernalisation treatment, plants from line 2_C1_BC2_S2_2 appeared to have some vernalisation response, with some plants bolting before or as WT following the cold treatment.

As all three LB individuals from line 164 originated from the same BC1 event only one of these individuals was analysed further; 164_E1_BC2_S2_3. All of the plants screened bolted significantly later than WT under non-vernalised conditions ($p=<0.001$; d.f.=19; l.s.d.=0.884); 7-12 days later, see figure 4.29. However after a

vernalisation treatment 15 out of 19 plants bolted as or before the latest WT plant to bolt. This suggests that line 164_E1_BC2_S2_3 has a vernalisation response. Although the result is not significant according to ANOVA (p=<0.001; d.f.=19; l.s.d.=0.917), the fact that the LB phenotype observed with no treatment is rescued by a vernalisation treatment, potentially means that the mutation causing the LB phenotype is present within an autonomous pathway gene as explained in section 1.4.4 and figure 1.7.

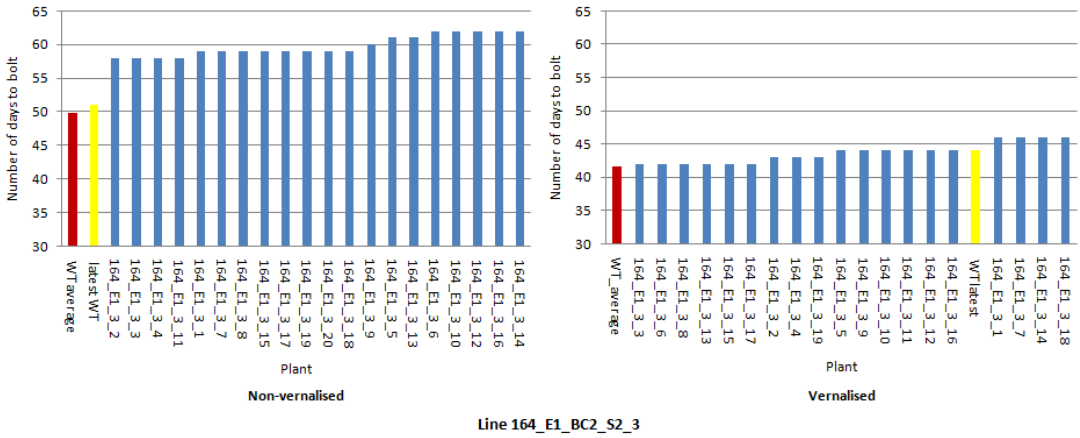


Figure 4.29 - Larissa line 164 BC2_S2 LB line scored for days to bolt with and without a vernalisation treatment. The red bar indicates the average number of days for WT plants to bolt, the yellow bar indicates the number of days that the latest WT plant took to bolt. All line 164 plants are later than WT without a vernalisation treatment, however the LB phenotype appears to have been rescued with a vernalisation treatment.

Perhaps the most drastic LB phenotype is apparent in line 307_B4_BC2_S2_4. Under both non-vernalised (p=<0.001; d.f.=19; l.s.d.=0.919) and vernalised (p=<0.001; d.f.=19; l.s.d.=1.069) conditions each of the 20 plants sown are significantly later than WT. There is a 17-27 day delay in bolting under non-vernalised conditions and an 11-18 day delay when the plants are vernalised, see figure 4.30 and 4.31. This suggests that the mutation causing the LB phenotype is not within a gene making up the autonomous pathway.

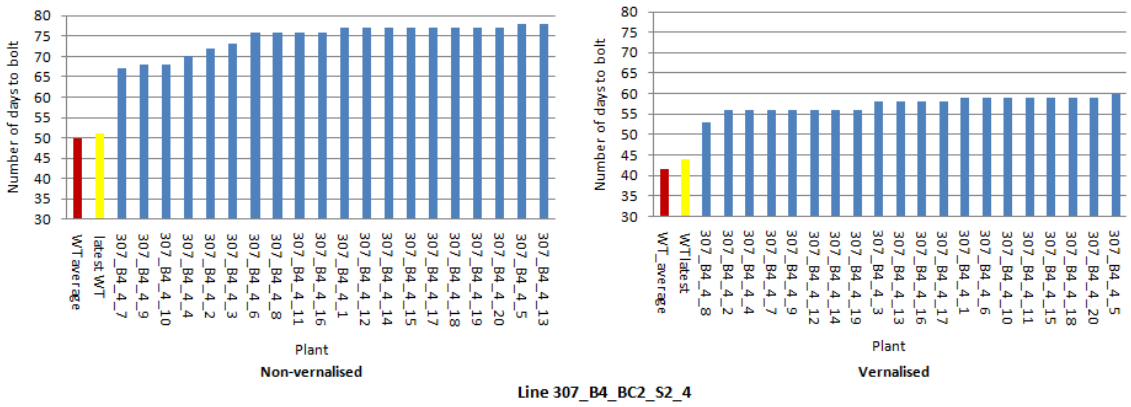


Figure 4.30 - Larissa line 307 BC2_S2 LB lines scored for days to bolt with and without a vernalisation treatment. The red bar indicates the average number of days for WT plants to bolt, the yellow bar indicates the number of days that the latest WT plant took to bolt. All line 307 plants are later than WT without or without a vernalisation treatment.

Line 307_B4_BC2_S2_4 appears to have the most robust LB phenotype of all the Larissa lines analysed. Furthermore, because its LB phenotype can be overcome with a vernalisation treatment, line 164_E1_BC2_S2_3 is also of great interest because the autonomous pathway genes targeted in this project all cause LB when mutated, a phenotype which can be overcome by a vernalisation treatment.

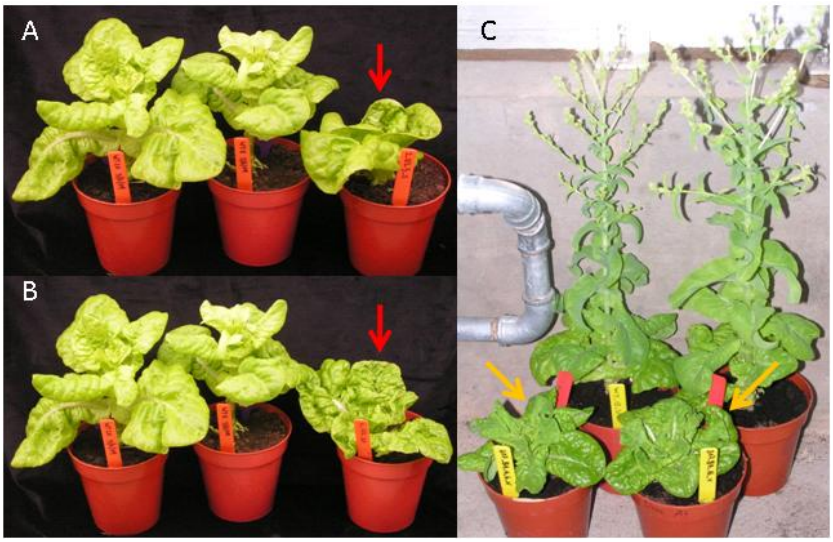


Figure 4.31 – Examples of homozygous mutant late-bolting Larissa lines. Lines 2_D3_BC2_S2_3 (A) and 2_G2_BC2_S2_1 (B) (highlighted with red arrows), both show no visual signs of bolting compared with the two WT plants in each photo which bolted 5-7 days earlier. Two plants representing line 307_B4_BC2_S2_4 (C) (highlighted with orange arrows), are beginning to show the first signs of bolting in comparison with the WT plants have already bolted and are close to flowering. All photographs shown are of non-vernalised plants.

4.5.2.6 Assessment of LB Larissa lines in the field

All of the Larissa experiments up to this point have been carried out in the glasshouse; the plants have been well maintained in this environment. To verify one of the agronomic aims of this project, that the LB lettuce identified should have an increased holding ability in the field, trials were set up to screen the LB Larissa lines 2, 164 and 307 in the field. Two trials were run at Wellesbourne, the plants for the first trial were transplanted in to the field in early spring (11th May 2010), a repeat of the trial was planted six weeks later in mid-summer (21st June 2010), to enable a comparison of the plants bolting phenotypes in the same location under different environmental conditions. A replicate of the second trial was planted at the farm of JE Piccaver Co., Spalding, Lincolnshire, to compare the bolting phenotype of the plants in two different locations.

Twenty plants representing each of the LB lines were grown alongside 40 WT plants, each was scored for the number of days to bolt. In this experiment, as opposed to previous Larissa trials in the glasshouse, no basal leaves were removed meaning that the rate at which each of the lines grew could be compared directly with WT. It also meant the LB phenotype could be verified in conditions more typical for lettuce production.

The data collected from each trial will be discussed individually. Figure 4.32 shows the data collected from the first field trial conducted at Wellesbourne. The 40 WT plants bolted after an average of 64.9 +/-0.38 days. Each of the LB Larissa lines reproduced the LB phenotype previously verified in the glasshouse, each of the lines bolted significantly later than WT ($p=<0.001$; d.f.=95; l.s.d.=0.525). However line 164_E1_3 was, on average only 2.25 days later to bolt than WT. The three lines representing the original LB M2 line 2; 2_C1_2, 2_D3_3 and 2_G2_1 were on

average, between 5 and 5.5 days later to bolt than WT. Interestingly, as seen in the glasshouse experiments line 307_B4_4 was the latest line to bolt, on average after 76.7 +/-0.47 days, 12 days later than WT.

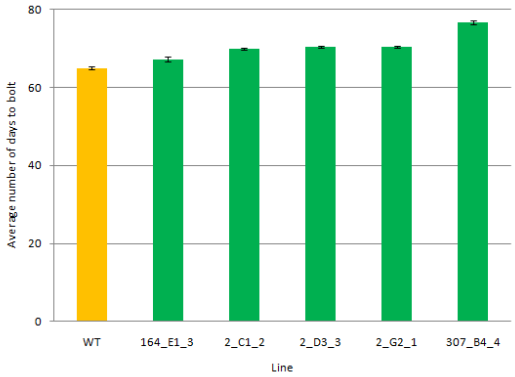


Figure 4.32 – Wellesbourne Field Trial I - Comparison of the average number of days to bolt for Larissa WT (orange) and the LB Larissa lines (green) in the field.

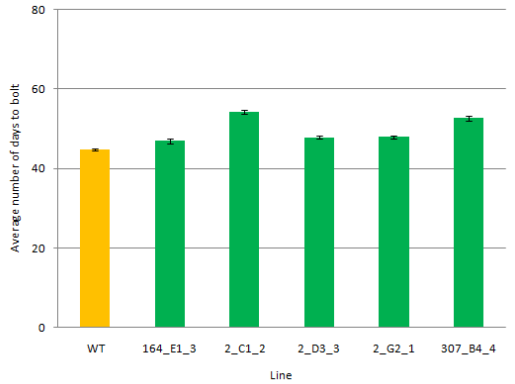


Figure 4.33 - Wellesbourne Field Trial II - Comparison of the average number of days to bolt for Larissa WT (orange) and the LB Larissa lines (green) in the field.

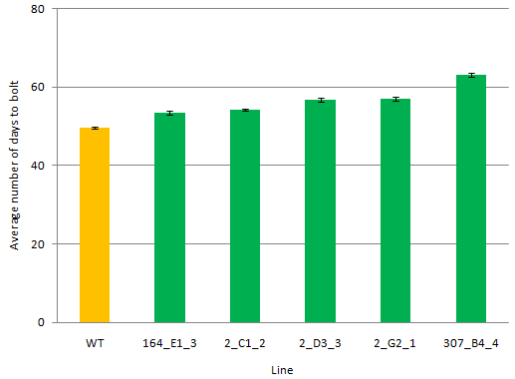


Figure 4.34 - Spalding Field Trial I - Comparison of the average number of days to bolt for Larissa WT (orange) and the LB Larissa lines (green) in the field.

The second field trial at Wellesbourne was transplanted into the field six weeks after the first and covered the mid-summer period. All the plants bolted earlier than those grown in the first trial. WT plants bolted, on average after 44.8 +/-0.2 days, 20.1 days earlier than the WT plants scored in the first trial. Bolting is initiated earlier under warmer conditions, the average daily maximum temperature for the first trial was 20.9°C compared to 24.1°C throughout the second trial (Horticultural Services, WHRI). This along with the longer photoperiods and higher irradiance (trial 1 – 15.30h-16.52h daylength, 377.68W/m^2 irradiance; trial 2 – 15.56h-16.52h daylength, 404.77W/m^2 irradiance) in June and July are reasons why the plants grown in the second trial bolted earlier. Figure 4.33 shows the data obtained from the second trial at Wellesbourne. All of the interesting LB lines screened bolted significantly later on average than WT ($p=<0.001$; d.f.=84; l.s.d.=0.94), as in the first trial line 164_E1_3 bolted on average only 2.45 days after WT. The plants from lines 2_D3_3 and 2_G2_1 bolted, on average just 3.1 and 3.15 days later than WT. Line 307 is still significantly later than WT (8 days later) in the second trial.

Of most interest from this data set, is that line 2_C1_2 is the latest line, on average to bolt, after 54.25 +/-0.55 days, 9.4 days later than WT. This line was on average only five days later than WT in the first field trial. This is an unexpected result as the other LB lines all display a similar delay in bolting compared to WT as seen in the first trial. The difference between line 2_C1_2 and the other lines derived from LB line 2 (2_D3_3 and 2_G2_1) could be due to a segregating background mutation that is present in line 2_C1_2 but not the other lines. This mutation could cause an additional effect on bolting time in response to one or more of the environmental variables between the two field trials at Wellesbourne. Figure 4.35 shows some examples of lettuce plants screened in the Wellesbourne field trial.

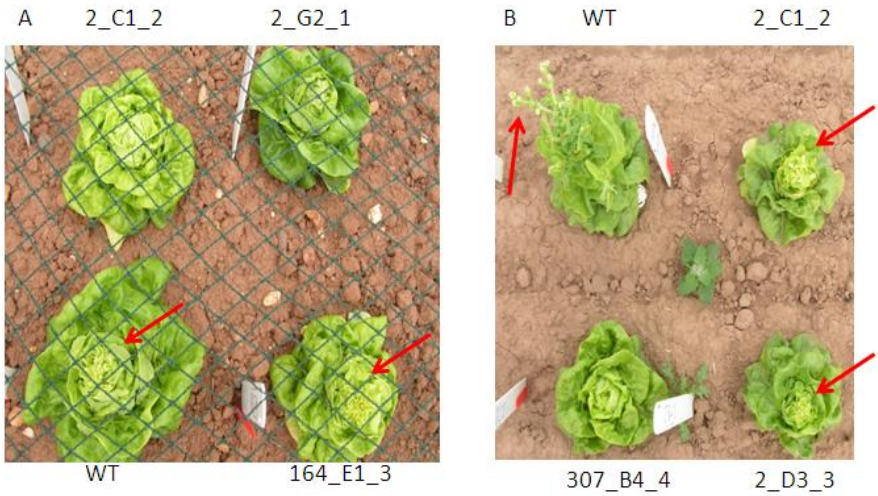


Figure 4.35 – Examples of LB Larissa lines in the field.
A – Wellesbourne Trial II – the red arrows indicate a WT and 164_E1_3 plant in which a bolt has initiated. The other two plants originating from Larissa line 2 are showing no sign of bolting. Photo taken 48 days after sowing.
B – Spalding Trial I – the red arrows indicate a WT plant which has started to flower and plants from 2_D3_3 and 2_C1_2 in which bolting has initiated. The other plant from Larissa line 307_B4_4 is showing no sign of bolting. Photo taken 61 days after sowing.
The phenotype and biomass of the mutant plants is similar to the WT plants grown.

A replicate of the Wellesbourne field trial was run in Lincolnshire, the data from which is shown in figure 4.34. Once again, the plants all bolted earlier than those in the first Wellesbourne trial, WT bolted after an average of 49.7 +/- 0.24 days. This was five days later than the WT plants scored in the second trial at Wellesbourne. All of the LB Larissa lines screened reproduced their LB phenotype, each line bolted significantly later than WT ($p=<0.001$; d.f.=82; l.s.d.=0.822).

Line 164_E1_3 bolted on average, 3.8 days later than WT, with lines 2_D3_3 and 2_G2_1 bolting on average 7.1 and 7.4 days later than WT respectively. Line 307_B4_4 was the latest line on average to bolt, after 63.2 +/-0.49 days, 13.5 days later than WT. Figure 4.44B shows some examples of lettuce plants screened in the Spalding field trial. Interestingly line 2_C1_2, which bolted latest compared to WT in the second Wellesbourne trial, bolted on average just 3.8 days later than WT in the Spalding trial, a similar delay to that recorded in the first Wellesbourne trial. Table

4.3 compares the average number of days to bolt for each of the lines recorded for the second Wellesbourne trial and the trial at Spalding.

Line	Average number of days to bolt	
	Wellesbourne Trial II	Spalding Trial I
WT	44.8	49.7
2_C1_2	54.25	53.5
2_D3_3	47.9	56.6
2_G2_1	47.95	57.1
164_E1_3	47.25	53.5
307_B4_4	52.25	63.2

Table 4.3 – Comparison of the average number of days to bolt for Larissa WT and the LB lines screened in the second field trial at Wellesbourne and the first field trial at Spalding.

In each case the average bolting times of the LB lines are all earlier in the trial which took place at Wellesbourne (between five and eleven days earlier), with the exception of line 2_C1_2 which bolts after roughly the same average number of days (54.25 and 53.5 days at Wellesbourne and Spalding respectively). After comparing the data collected from all of the LB lines in each of the three trials with a two-way ANOVA, it is interesting to see that 2_C1_2 plants grown in the second Wellesbourne trial and the Spalding trial is the only line not showing a significant difference ($p=<0.001$, d.f.=10; l.s.d.=1.333) as shown in figure 4.36. As previously discussed a possible reason for the phenotype seen in line 2_C1_2 in the second trial may be that in this line a second site mutation is present within a gene regulating plant development in response to the environment, which is absent from lines 2_D3_3 and 2_G2_1.

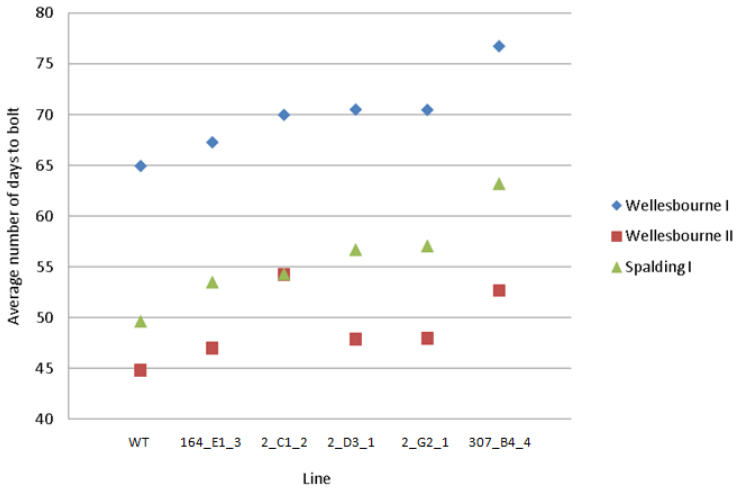


Figure 4.36 – Average number of days to bolt for each of the LB Larissa lines in each of the three field trials. Each line bolted earliest in the second Wellesbourne trial, except line 2_C1_2 which bolting at a similar time in the Spalding trial.

The average maximum temperature recorded at Wellesbourne for the second trial was 24.1 °C compared to 23.0 °C at Spalding, however the maximum temperature in Wellesbourne was over 25 °C on twelve of the days throughout the trial, compared with only six in Spalding. The weather system affecting the Spalding trial also involves cooling winds which come from the nearby coastline, Wellesbourne being inland is not affected by such a system. As would be expected the photoperiod recorded throughout both trials did not vary significantly between the two sites (Wellesbourne 15.56h-16.52h; Spalding 16h-16.58h, Met office), suggesting that photoperiod was not an important factor in the bolting phenotype of 2_C2_2, see table 4.4.

Trial	Ave. Max. Temp. (°C)	Ave. Min. Temp. (°C)	Temp. range (°C)	P'period range (h)	Ave. p'period (h)	Irradiance (W/m^{2})
W'bourne I	20.9	8.7	-1-27.9	15.30-16.52	16.20	377.68
W'bourne II	24.1	12.3	8.1-27.9	15.56-16.52	16.30	404.77
Spalding I	23.0	13	9-23	16-16.58	16.34	N/A

Table 4.4 – Weather conditions recorded at each field trial site

In summary, each of the LB Larissa lines identified from screens which have taken place in the glasshouse, replicate the LB phenotype in the field. Each line looks like

WT and grows at a rate comparable to WT (although no accurate measure of this was made), these are important aspects for the grower. Lines 164_E1_4, 2_D3_3 and 2_G2_1 bolted on average just a few days after WT in each of the three trials, compared with 307_B4_4 which bolted significantly later than WT in all three trials, replicating the extremely LB phenotype seen in the glasshouse. Line 2_C1_2 produced the most interesting data, bolting, on average four to five days later than WT in the first Wellesbourne trial and in the Spalding trial, but under the higher temperatures of the second Wellesbourne trial, bolted on average 9.4 days later than WT and later than all of the other lines in that trial.

4.5.2.7 Assessment of the temperature on the LB Larissa lines

As previously discussed temperature has a significant effect on the time at which lettuce plants bolt. To investigate the effect of temperature on the bolting time of the three LB Larissa lines; 2, 164 and 307, four plants representing each line and WT were grown in growth cabinets at constant temperatures of 18 °C, 21 °C, 23 °C and 25 °C in 16h photoperiods. 18 °C, 21 °C and 23 °C are the average temperatures recorded through the spring and summer months in the UK. 25 °C was used to see if the plants maintained their LB phenotype under a higher temperature, with the current trend for warmer summers, more extreme growing conditions need to be examined. Figure 4.37 summarises the average number of days to bolt per line under each temperature.

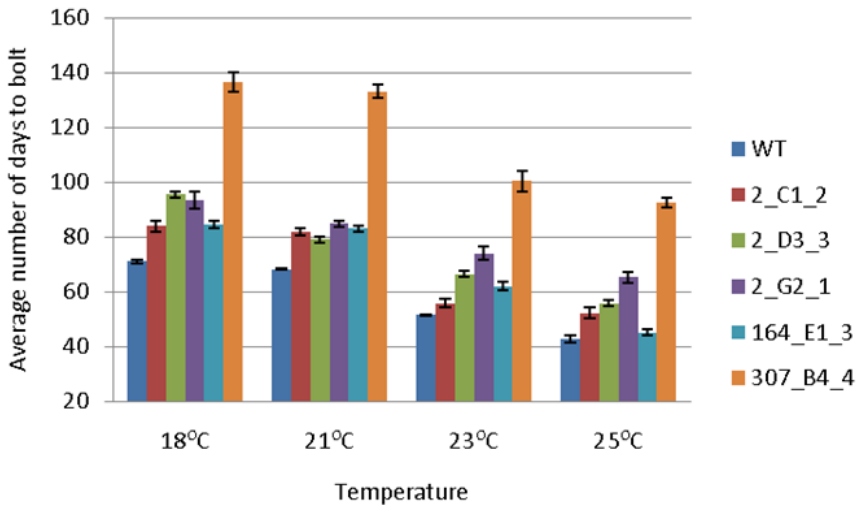


Figure 4.37 – Average number of days to bolt for Larissa WT and LB Larissa lines under different temperatures

Each line bolts slightly earlier with each increase in temperature. Each of the lines bolt significantly later on average than WT at both 18 °C (p=<0.001, d.f.=15, l.s.d.=8.08) and at 21 °C (p=<0.001, d.f.=15, l.s.d.=4.393). The LB lines do however show a bigger decrease in the number of days to bolt compared with WT plants between the temperatures of 18 °C and 21 °C. Each line, including WT bolts much earlier at 23 °C; all lines except 2_C1_2 (which bolts 4.25 days later), bolt significantly later on average than WT (p=<0.001; d.f.=15; l.s.d.=6.39) at this temperature. At 25 °C the plants all bolt earlier still, however all of the lines bolt significantly later on average than WT (p=<0.001, d.f.=15; l.s.d.=7.39). The earlier bolting at higher temperatures is consistent with research conducted on *Arabidopsis* which has identified an ambient temperature flowering pathway (Blázquez *et al*., 2003). *Arabidopsis* plants grown at 23 °C were found to express higher levels of *FT* and *SOC1* compared to plants grown at 16 °C this resulted in earlier flowering.

The LB phenotype of line 164_E1_3 appears to be temperature dependent; its LB phenotype is strong at 18-21 °C, at 23 °C there is still a ten day delay compared to

WT. However at 25 °C line 164_E1_3 is only 2.25 days later than WT, suggesting that this line struggles to maintain its LB phenotype at more extreme temperatures. Interestingly, line 2_C1_2 which did not show a significantly LB phenotype at 23 °C, does so at 25 °C. Line 2_C1_2, which showed the unusual bolting phenotype in the field trials, again shows a different bolting response compared to the other two lines, 2_D3_3 and 2_G2_1, which are also derived from the original LB line 2. The delay in bolting between lines 2_D3_3 and 2_G2_1 and WT is fairly consistent, particularly between 21 °C and 25 °C, see figure 4.38. The delay observed in line 2_C1_2, however, shows a different temperature response and appears more variable, becoming greater as the temperature increases from 23 °C to 25 °C (orange arrow, figure 4.48). The bolting of this line is clearly responding to temperature in a different way to the other two lines derived from the original LB line 2, and as proposed earlier, this could be due to a segregating second site mutation.

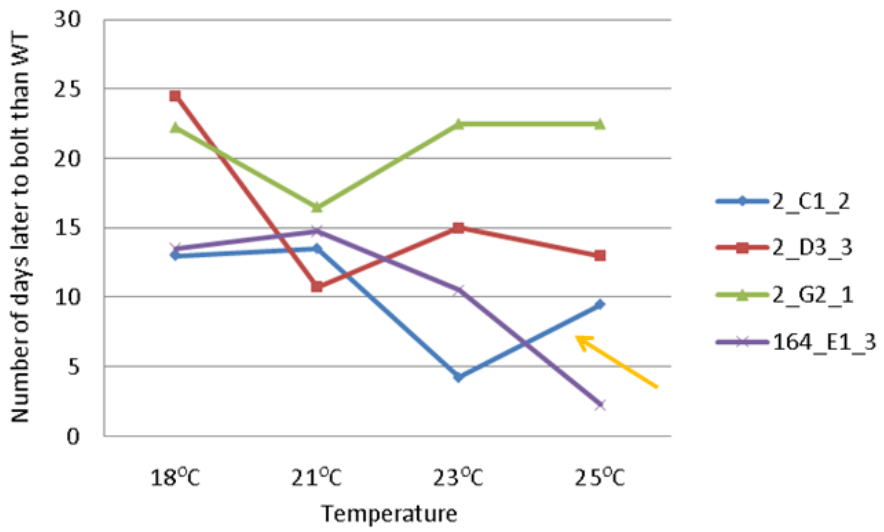


Figure 4.38 – The number of days later to bolt for each of the LB Larissa lines compared to WT. The LB phenotype of line 2_C1_2 appears to improve at higher temperatures.

It is encouraging to see that all three lines (2, 164 and 307) bolt later than WT even at 25°C, suggesting the lines developed have a robust phenotype in different temperatures. Line 307, which has consistently bolted later than WT in all the trials

once again is the latest line to bolt, 49 and 49.75 days later on average than WT at 23°C and 25°C respectively.

Figure 4.39 shows the consistently LB lines 2_D3_3, 2_G2_1 and 307_B4_4 compared to WT at some of the temperatures tested.

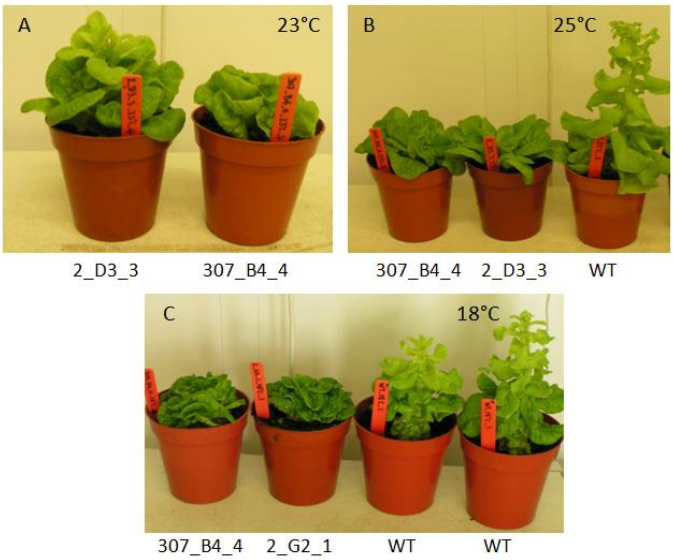


Figure 4.39 – Examples of plants grown to assess the phenotype of the LB Larissa lines under different temperatures. (A) – Line 307_B4_4 is yet to bolt at 23°C after 77 days even though another LB line 2_D3_3 has bolted . (B) – Lines 307_B4_4 and 2_D3_3 compared to WT at 25°C after 50 days, WT has bolted whereas the two LB lines have not. (C) – Lines 307_B4_4 and 2_G2_1 compared to WT at 18°C, after 83 days, WT plants have bolted whereas the two LB have not.

4.5.3 Summary

The Larissa EMS population has produced some interesting results from what was a relatively small population created to test the feasibility and efficiency of mutagenesis protocols. Seven interesting LB lines were identified from an initial screen of the M2 and M3 seed made available, the robustness of the phenotype was confirmed before the plants were BC to WT, to reduce the number of background EMS mutations, the plants were then selfed to obtain homozygous LB mutant lines. Three lines; line 2, 164 and 307 all produced progeny with a significant and reproducible LB phenotype. Furthermore the three lines have maintained their LB

phenotype when grown in the field at different times of the year and in different locations and under increased temperatures.

It appears from the phenotypes of the segregating plants screened that the mutation causing the LB phenotype in each line is dominant. It would have been interesting to look at the make-up of all the Larissa lines which were screened in more detail, to try to identify lines containing plants containing recessive mutations, e.g. line 72 includes one significantly LB plant out of five M2 plants screened and line 90 includes two significantly LB plants out of seven M2 plants screened, a LB mutation caused by a recessive mutation may be segregating at this point. The lack of any M1 data for the Larissa population, and the fact that it was only a trial for the larger Saladin EMS population, meant that time was invested in working with the most obvious LB lines.

4.6 Results and Discussion – Saladin Population

4.6.1 Screening of L. sativa cv. Saladin EMS mutagenised population for late-bolting lines

The second EMS mutagenised population is in the *L. sativa* cv. Saladin background. Saladin is a crisphead lettuce variety which is slower to mature, and therefore requires more maintenance than Larissa. To allow bolting to be accurately scored the heads of the Saladin plants were regularly cut transversely, this meant that developing floral parts could be identified easily and prevented the developing bolt from rotting, see figure 4.5. Saladin will only bolt under natural LD photoperiods, meaning that only one generation of plants were able to be grown per year.

4.6.1.1 'Test' population – M2 screen

The 'test' population was derived from the seed batch used to optimise the EMS treatment (concentration, time etc.) before mutagenising the 'main' batch of seed. This population was produced one year in advance of the 'main' population and was comprised of 125 M2 Saladin lines. M1 seed was not available from this population for this project. Four plants from each M2 line were grown in the field and were screened for number of days to bolt under the protection of haygrove tunnels, see figure 4.40. As with the Larissa EMS population that was screened, a number of plants displayed phenotypic imperfections giving an indication that the EMS treatment had been successful.

Figure 4.40 - Saladin TILLING Population – Plants in the field under the haygrove polytunnels (A), Saladin plant bolting in the field (B).

Of the 125 lines grown, 107 produced at least one plant which survived to bolt. A large number of plants were lost in this trial due to a wire worm infestation. As with the Larissa EMS population a large variation in days to bolt was observed, the earliest plants bolted after 85 days and the latest after 194 days; a range of 109 days. WT Saladin plants bolted after an average of 124.8 +/-0.75 days, the spread of bolting times recorded for the 'test' Saladin population is shown in see figure 4.41.

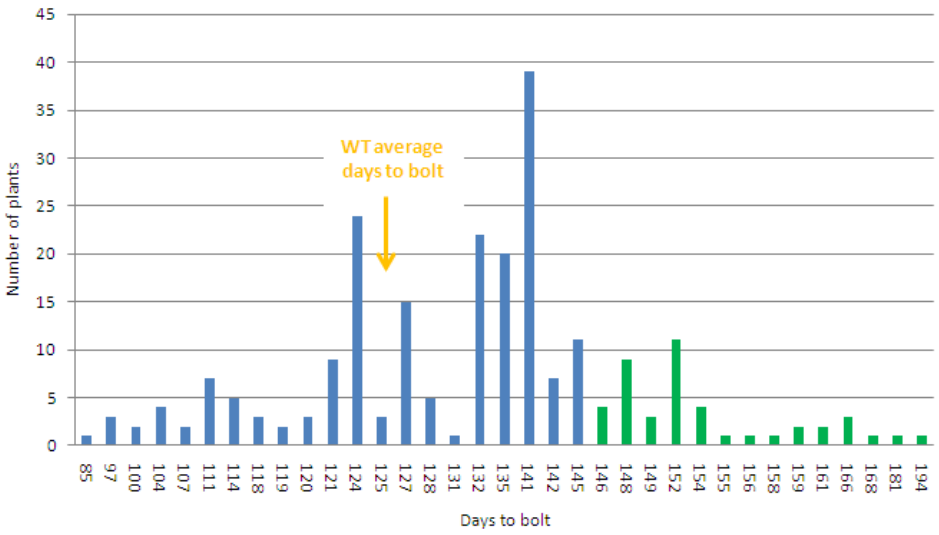


Figure 4.41 – Saladin 'test' population - M2 generation. The graph illustrates the range of bolting times for the population. The green bars indicate LB plants and the orange arrow indicates the average number of days to bolt for WT plants.

Figure 4.42 shows an example of Saladin plants at different developmental stages in the field, including some which have bolted.

Figure 4.42 – Example of Saladin plants bolting in the field.

Plants bolting with an average of 141.3 days or more were significantly later than WT (p=<0.001, d.f.=124, l.s.d.=16.3). This included 33 Saladin lines; which were too many to score and screen in the glasshouse, a compromise was made whereby the plants bolting after 146 days or more (~20 % of population), two weeks later than WT were scored as LB. 42 plants from 15 different lines bolted after 146 days or more, these are highlighted in green in figure 4.51.

A Q-Q plot analysis was calculated to further test the likelihood that the plants chosen were significantly LB compared with the rest of the population screened. The average number of days to bolt for each line (marked with a cross) are within the 90 % confidence limits (marked with red lines) indicating that the data fits a Gaussian distribution. The Q-Q plot illustrated in figure 4.43 was used to calculate that there is only a 12.5 % chance that the 15 lines classified as LB would bolt with an average of 146 days or more. This suggests that the lines chosen are significantly different from the rest of the population and provides confidence that the LB phenotype observed is less likely to be caused by natural variation this indicates that something else is contributing to the bolting phenotypes observed.

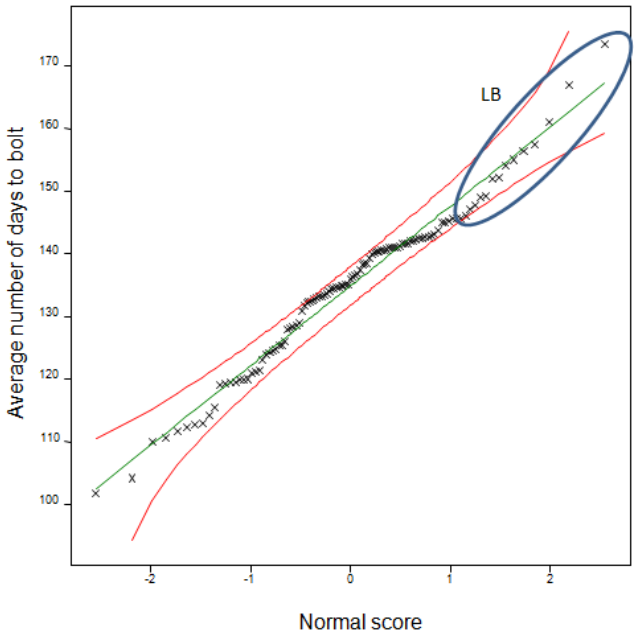


Figure 4.43 – Normal Q-Q probability plot for trt (+ 90 % confidence limits) for Saladin 'test' EMS population. The skewed red lines represents the 90 % confidence level that the data fits a Gaussian distribution, the lines highlighted in blue are LB lines, these phenotypes are less likely to be caused by natural variation, indicating something else is causing the bolting phenotype observed.

The 15 lines bolting significantly later on average than WT are highlighted in green in figure 4.44. The cut off for LB was difficult to assign accurately as a number of plants were infected by a wire worm infestation; these plants took longer to recover than non-infected plants and as a result could be deemed to be LB due to the effects of the wire worm and not the EMS treatment. Plants which were infected were noted and the data was analysed with this in mind. Of the 15 LB lines, all of the plants screened for 13 lines were LB, the other two; lines 75 and 126 each contained two individual LB plants. Seven of the LB lines contained plants that were infected by wire worm and therefore these plants were not considered as LB in case the effects of the infestation had slowed their growth and resulted in the LB phenotype. The plants from the eight remaining lines did appear to be LB due to the effects of the EMS treatment; they included line 75, where two of the four plants scored were LB, while all of the plants making up the other seven lines; lines 10, 38, 88, 100, 117, 131 and 138, were LB (highlighted in figure 4.54 with red arrows). A further 12 Saladin

lines, which bolted on average earlier than 146 days and were not affected by the wire worm infestation contained one plant that bolted significantly later than WT, (highlighted in brown) these plants are also of interest as they may be carrying a recessive mutation that is causing the LB phenotype.

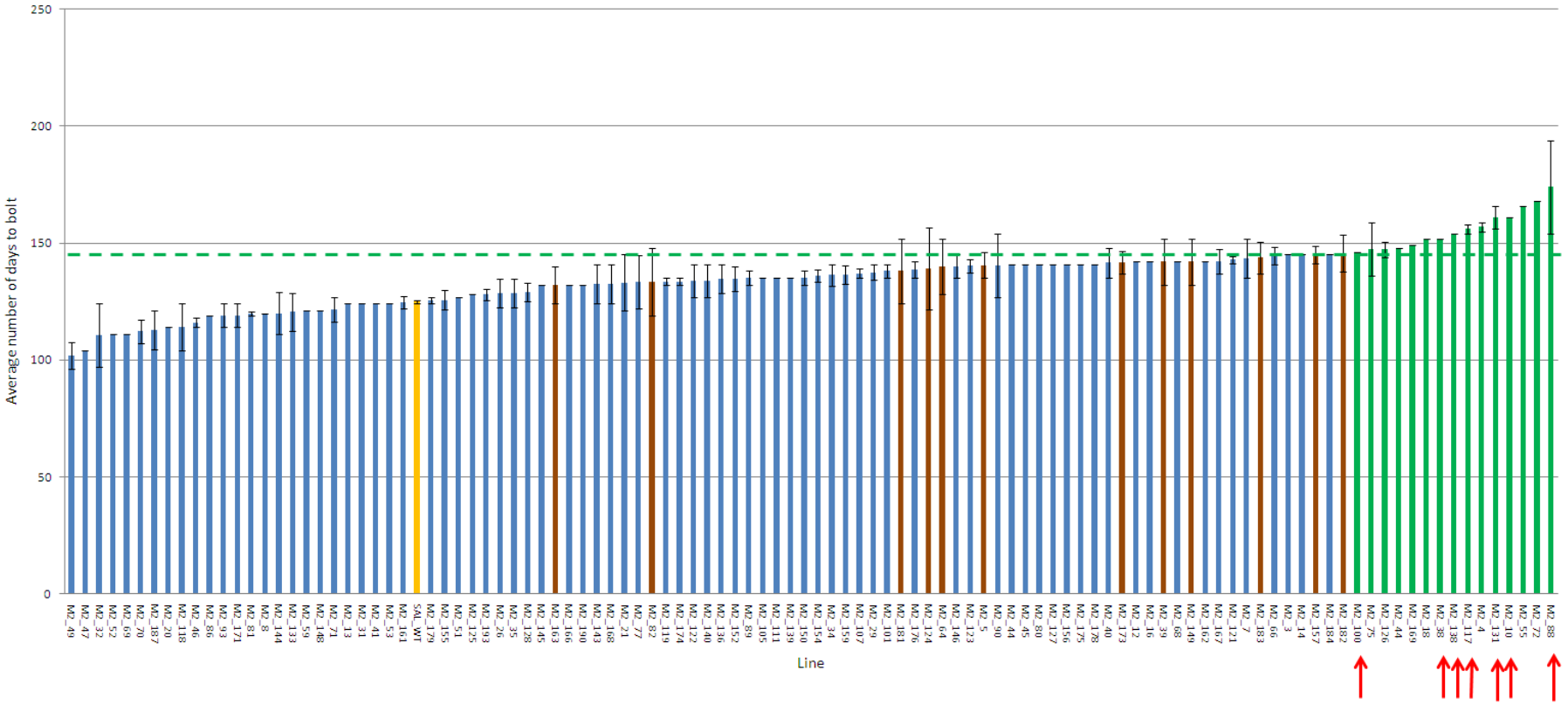


Figure 4.44 – Saladin 'test' population M2 - average number of days to bolt per line. WT is highlighted in orange, lines containing plants which all bolted later than 146 days were classified as LB and are highlighted in green, lines containing between one and three plants which bolted after 146 days are highlighted in brown. Some plants were affected by a wire worm infestation, which potentially caused the LB phenotype observed. The seven LB lines in which all plants scored were LB and were not affected by a wire worm infestation are highlighted with red arrows. The green dotted line indicates the cut off point for a plant to be scored as LB.

4.6.1.2 'Test' population – repeat of M2 screen in glasshouse

Four M2 plants from each of the eight LB lines and the twelve lines containing individual LB plants were grown under a natural LD photoperiod in the glasshouse to verify the LB phenotype and to BC to WT, as described in sections 2.1.2.1 and 2.1.2.3. Not all of the plants survived to bolting in the glasshouse because Saladin lettuce plants take a relatively long time to mature, the plants get large when grown in the field, see figure 4.42, but their growth is restricted in pots in the glasshouse. Some plants struggle to develop with limited root space, this along with the risk of disease created by large leaves meant that more plant losses were incurred than were seen with the smaller, more compact Larissa plants when grown in the glasshouse. Interestingly all of the plants that bolted did so earlier than the plants grown in the field; WT plants bolted after an average of 102 +/-2 days in the glasshouse compared with 124.8 +/-0.75 days in the field. Lines bolting on average after 121 days or more are significantly different from WT ($p=<0.001$, d.f.=28, l.s.d.=18.99). Only four lines; lines 38, 39, 75 and 131 fit this criteria, which for the purposes of this work seemed overly stringent; therefore plants which bolted after 112 days or later, ten days later than WT were classified as LB.

Of the eight LB lines rescored, plants representing three of the lines; line 38, 131 and 138, reproduced the LB phenotype previously seen in the field, whereby all of the plants bolted ten days or later than WT. Line 75, which contained two LB plants of the four sown in the field trial, produced plants which all bolted after 102 days or more in the glasshouse. Two of the eight lines; line 88 and 117 had no LB phenotype in the glasshouse. The plants from the two other LB lines; line 10 and 100 died before bolting.

Of the 12 lines which contained one LB plant in the field screen, line 39 produced plants which all bolted after 112 days or more in the glasshouse. A further three lines; line 64, 124 and 157 contained some plants which bolted after or later than 112 days. Four of the lines; 163, 173, 182 and 183 did not contain any plants with a LB phenotype and the plants from the final four lines all died before bolting. Figure 4.45, highlights the number of days to bolt per line against WT, the orange dotted line highlights the average number of days to bolt for WT plants and the green dotted line indicates the number of days to bolt required for a plant to be scored LB.

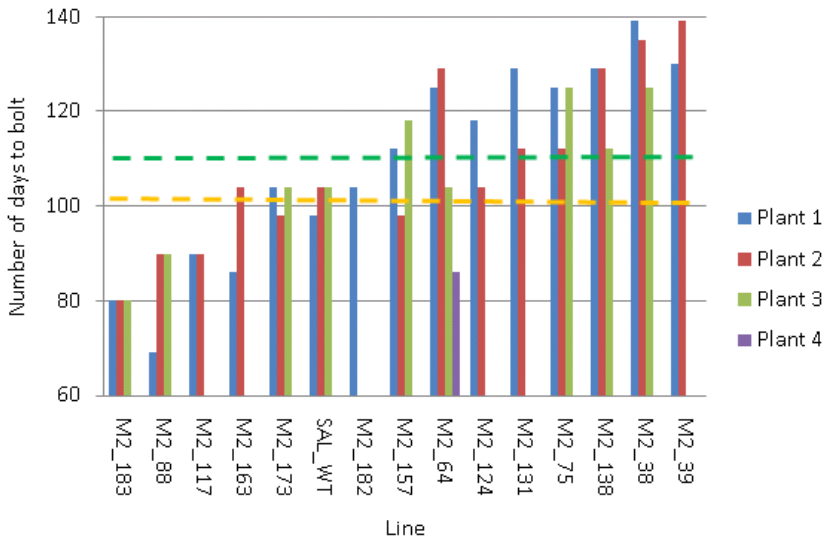


Figure 4.45 – Rescoring of M2 Saladin lines in the glasshouse which were previously identified as LB in the field. Four plants from each LB line were scored for number of days to bolt. The average number of days to bolt for WT is highlighted with an orange dotted line. Plants bolting after 112 days (highlighted with a green dotted line) or more were scored as LB. Five lines have been confirmed as LB, with a further three containing at least one LB plant

Therefore after two trials, one in the field and one in the glasshouse, three Saladin lines, lines 38, 131 and 138, with a reproducible LB phenotype, where all plants scored have bolted significantly later than WT, have been identified from the 'test' population. A further two lines; lines 39 and 75 have produced plants that are all LB in the glasshouse, and individual plants that bolted significantly later than WT in the field trial. Three lines; lines 64, 124 and 157 were represented by individual LB plants in both the field and the glasshouse trials.

4.6.1.3 ‘Test’ population – repeat of M2 screen in the field

Due to the problems that occurred with the wire worm infestation in the original screen, where enough seed was available four plants from each of the M2 lines from the ‘test’ population were re-grown in the field the following summer in 2008. In total 103 of the 125 lines were re-screened, WT plants bolted after an average of 125 +/-0.6 days similar to the previous year. Plants bolting after 136.4 days or more were significantly later than WT (p=<0.001; d.f.=213; l.s.d.=11.417).

It was decided to compromise slightly on this figure, to bring this data into line with the first screen. Plants which bolted after 138 days or more, roughly two weeks later than WT were scored as LB this amounted to about 20 % of all the plants screened, as with the first field trial screen. The spread of days to bolt for the population ranged from 99-168 days, (69 days), this is considerably less than the range observed in the first field trial, of 109 days. This could be due to a number of reasons, firstly the wire worm may have had a more significant effect on all of the plants screened in the first trial, this may have caused the development of the plants to be delayed. Another potential reason for the difference may have been that the plants grown in the second trial were maintained at a much higher level, basal leaves were removed earlier in the plant’s development and the heads were cut on a more regular basis meaning that scoring was not only easier, but more accurate. Weather conditions were probably not significant, neither summer was particularly warm and the plants making up both trials were both transplanted in the first week of May (2007 and 2008). However, the average number of days to bolt for the WT plants screened in each trial did not vary significantly (124.8 and 125 days respectively).

Nine of the lines screened in the second field trial contained plants which were LB, these are highlighted in green in figure 4.47. The LB phenotype in four of these

lines, lines 38, 39, 100 and 138, indicated in figure 4.57 with red arrows, had previously been observed in both the first field trial and the glasshouse screen. A further line; line 5 contained a single plant that bolted later than WT in the original field screen; the line bolted after an average of 140.5 +/-5.5 days compared to WT, which bolted after 124.8 +/-0.75 days. The other four lines; lines 47, 51, 56 and 145 were not identified as LB in the first screen, although only one plant per line survived to bolt in the first screen for lines 47, 51 and 145, and each of the plants representing line 56 died before they bolted. Figure 4.46 shows a Q-Q plot constructed to further support the significance of the nine LB lines (highlighted in blue) identified from this screen. The lines show a significant variation in bolting time from the rest of the population; it has been calculated from this plot that there is only a 7.5 % chance that these lines would bolt after 138 days or more, suggesting the phenotype may be caused by the EMS treatment, but is unlikely to be caused by natural variation within this population.

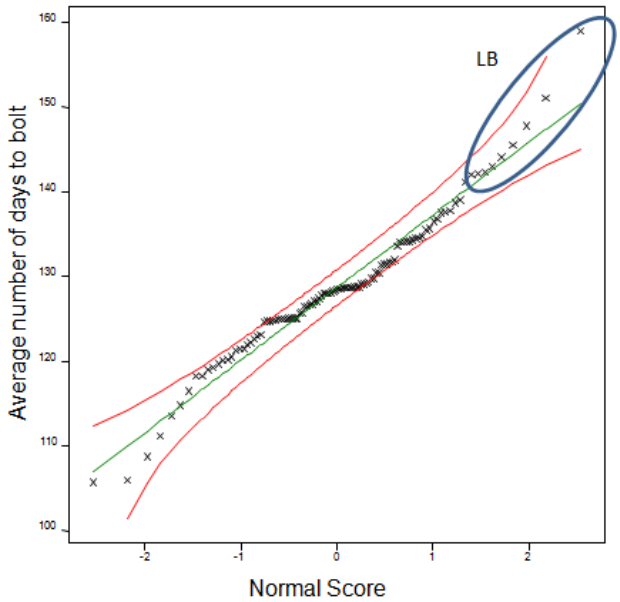


Figure 4.46 – Normal Q-Q probability plot (+ 90 % confidence limits) for Saladin EMS 'test' population – second field trial screen. The skewed red lines represents the 90 % confidence level that the data fits a Gaussian distribution, the lines highlighted in blue are LB lines, these phenotypes are less likely to be caused by natural variation, indicating something else is causing the bolting phenotype observed.

A further 26 lines, highlighted in brown in figure 4.47, contained one or two plants bolting after 138 days or more during the second screen, of these, two lines; lines 88 and 117 were LB lines when screened in the field previously, however they did not replicate the LB phenotype in the glasshouse. Line 75, contained an individual LB plant in this screen, this line was previously identified as a line containing one LB individual. Of the other 23 lines, none had been identified in the original screen as containing LB plants; 16 of the 23 lines however, had been severely affected by the wire worm infestation in that screen; the remaining seven lines all contained individual plants bolting later than WT, but not as late as the cut off of 146 days set for LB plants for the trial. These 23 lines were not followed up in this project, due to time restrictions.

Of the other lines previously identified as being of interest; lines 64, 124 and 131, which were LB in the first field trial and the glasshouse screen were not re-sown due to a shortage of seed. Line 157, in which all plants were LB in the original field trial and some plants were LB in the glasshouse, contained no LB individuals in this repeat trial. Lines 163, 173, 182 and 183, all of which contained some plants bolting later than WT in the first field trial, but no LB plants in the glasshouse, were re-scored, but no plants from any of the lines were LB, therefore confirming them as uninteresting.

This data along with the original field trial and the glasshouse trial means that LB Saladin M2 lines have been identified which are consistently reproducing a LB phenotype, these lines are summarised in table 4.5 and were taken forward for BC to WT, see section 2.2.16 for details.

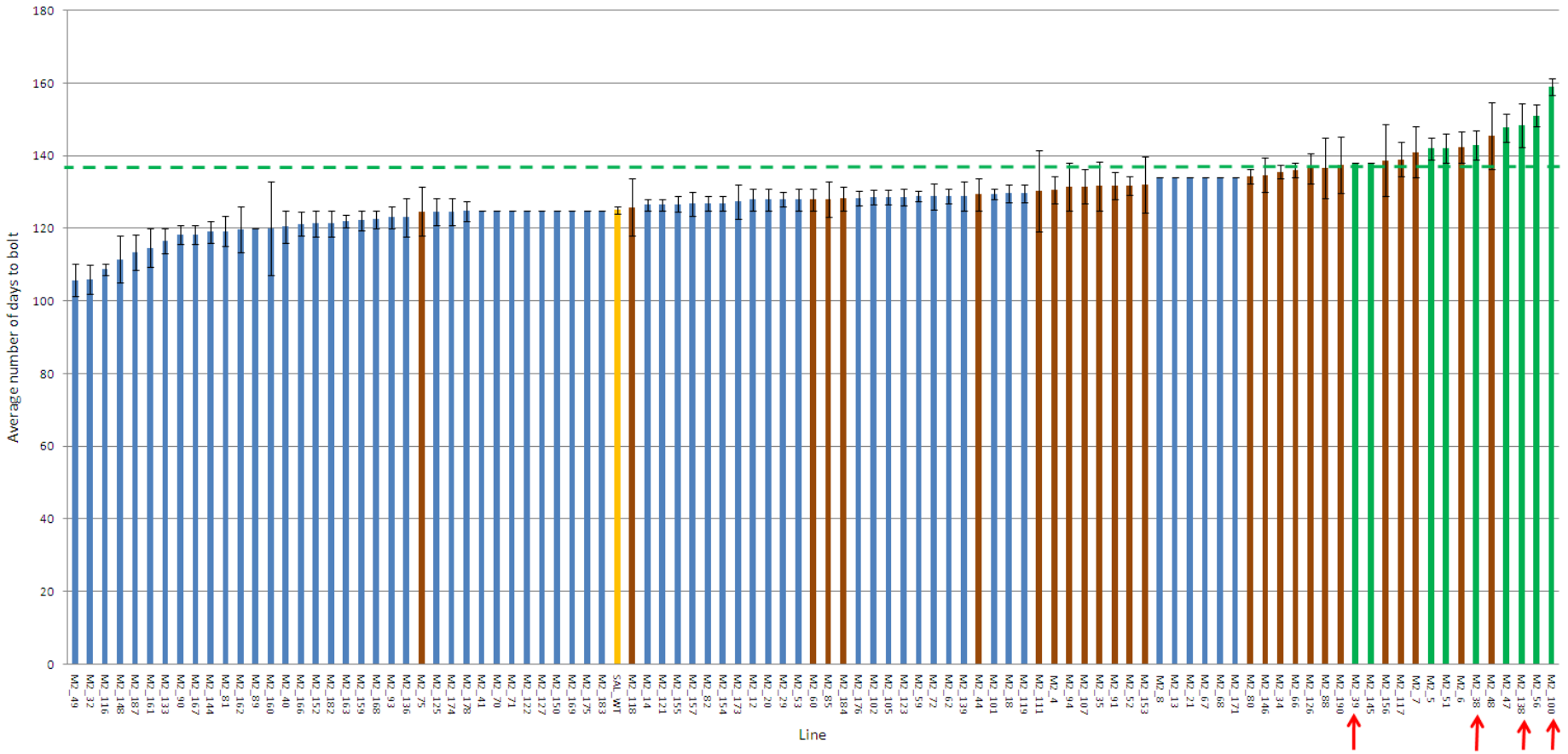


Figure 4.47 – Saladin 'test' population repeat screen – average number of days to bolt per line. WT is indicated by orange bar with an average of 125 days to bolt. Lines containing one or two plants with a LB phenotype, ≥138 days (highlighted with a green dotted line), are highlighted in brown, those lines in which all plants display a LB phenotype are highlighted in green. Lines in which all plants are LB in the first field screen and in this screen are highlighted with red arrow.

4.6.1.4 'Test' population – Screen of residual seed

Some of the M1 plants making up the M2 'test' population which was screened, particularly those which bolted later than WT, set very few seed before being collected. 29 M1 lines produced between one and six seed each, these individuals were not screened in the original trial in the field with the other 125 lines. Where seed was available, four plants representing each of the 29 lines were sown and scored for days to bolt in the glasshouse under a natural LD photoperiod. Seed representing eight of the lines did not germinate or the plants died before bolting. WT plants bolted on average after 102 +/-2 days, highlighted in orange on figure 4.48. To be consistent, plants which bolted after 112 days or more were classified as LB; (ANOVA suggests that plants bolting after 113 days or more are significantly different from WT, p=<0.001; d.f.=13; l.s.d.=10.998). Therefore 12 of the lines contained plants which all bolted significantly later than WT, and a further four lines contained one plant which bolted 10 days or later, than WT, highlighted in green and brown respectively in figure 4.48. These 16 LB lines are summarised in table 4.5 and were used to BC to WT, see section 2.2.16.

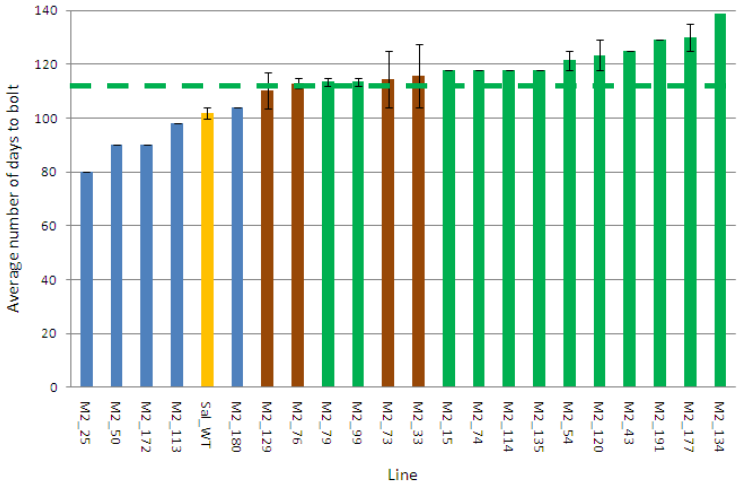


Figure 4.48 – Residual M2 plants from the Saladin 'test' population. The average number of days to bolt for WT plants are highlighted in orange, lines containing one plant with a LB phenotype are highlighted in brown and lines in which all plants displayed a LB phenotype are highlighted in green. Plants which bolt after 112 days or more were classified as LB (highlighted with green dotted line).

Table 4.5 summarises the findings from the Saladin 'test' population experiments described in sections 4.6.1.1-4.6.1.4; lines 38, 131 and 138 appear to display the most robust LB phenotype; every plant scored in both field trials and in the glasshouse trial were LB compared with WT. 12 residual lines also produced LB plants, however these were only scored once and some lines were comprised of only one plant. Other interesting lines appear to be line 39; only one plant scored from three trials did not display a significant LB bolting phenotype. Lines 88 and 117 contained LB plants in both of the field trials, but bolted as WT in the glasshouse and so were not BC to WT.

Test population				Test population – residual seed			
Line	**Field I**	**Glasshouse**	**Field II**	**Line**	**Field I**	**Glasshouse**	**Field II**
M2_10		Plants died	Not sown	**M2_79**	N/A		N/A
M2_38				**M2_99**	N/A		N/A
M2_88		No LB		**M2_15**	N/A		N/A
M2_100		Plants died		**M2_120**	N/A		N/A
M2_117		No LB		**M2_54**	N/A		N/A
M2_131			Not sown	**M2_135**	N/A		N/A
M2_138				**M2_114**	N/A		N/A
M2_5		Plants died		**M2_74**	N/A		N/A
M2_39				**M2_177**	N/A		N/A
M2_75				**M2_43**	N/A		N/A
M2_64			Not sown	**M2_191**	N/A		N/A
M2_124			Not sown	**M2_134**	N/A		N/A
M2_157			No LB	**M2_33**	N/A		N/A
M2_163		No LB	No LB	**M2_76**	N/A		N/A
M2_173		No LB	No LB	**M2_73**	N/A		N/A
M2_182		No LB	No LB	**M2_129**	N/A		N/A
M2_183		No LB	No LB				
M2_149		No seed	No seed				
M2_181		No seed	No seed				
M2_82		No seed	No LB				

Table 4.5 – Identification of LB lines from the M2 Saladin 'test' EMS population. Where seed was available, the lines were scored on three occasions, twice in the field and once in the glasshouse. The residual seed making up the population was only scored in the glasshouse. Green boxes indicate lines in which all plants bolted later than WT, brown indicates lines containing one or two plants bolting later than WT.

4.6.1.5 'Test' population – Identification of homozygous mutant lines

The LB plants identified in the glasshouse (sections 4.6.1.2 and 4.6.1.4) which confirmed the LB phenotype seen in the field were used to BC to WT. As with the

Larissa population, BC was performed in an attempt to remove undesirable EMS induced background mutations from the LB lines; this will be followed by self fertilisation to create homozygous mutant lines with a LB phenotype. LB plants from the 24 lines which displayed a LB phenotype in the glasshouse screen, highlighted green and brown in table 4.5, were BC to WT. As many BC events as possible per plant were performed. In general Saladin plants were far more difficult to BC than Larissa. Saladin flowers were much smaller and produced less pollen, if insufficient natural daylight was present when the flowers opened in the morning then the flowers did not open fully and produced less pollen than normal, which made back-crossing extremely difficult.

Of the three lines which were 100 % LB under all conditions (38, 131 and 138); lines 38 and 131 both successfully produced back-cross seed (BC1 seed), however, line 138 plants displayed an abnormal phenotype; the leaves on plants in this line were smaller than in WT, and BC was also unsuccessful, with limited seed being produced. Due to the phenotypic imperfections and the problems encountered with BC, line 138 was omitted from further experiments. Lines 39 and 75, which contained some LB plants in all three trials were also BC to WT, line 75 was successful, but no seed was produced from line 39. The two lines which contained one or two plants displaying a LB phenotype; lines 64 and 124 both successfully set BC1 seed.

The twelve extra LB lines identified from the residual seed were also BC to WT. Lines 15, 43, and 99 produced plenty of BC1 seed, however lines 54, 74, 79, 114, 120, 134, 135, 177 and 191 either had phenotypic imperfections or produced no seed. The individual LB plants representing the remaining four lines from the residual seed

which contained one or two LB plants in the glasshouse trial were also BC to WT, lines 73, 76 and 129 produced seed but line 33 did not produce any BC1 seed.

Overall BC1 seed was obtained from five lines where the plants all displayed a LB phenotype and from six lines where one or two of the plants sown were LB compared to WT. Relatively small amounts of BC1 seed was collected from the Saladin 'test' population, this may be due to the difficulties experienced when BC or due to the EMS treatment. Where possible six seed from at least one back cross event from at least one BC1 plant was sown and scored for days to bolt under a natural LD photoperiod. Around 20 % of the BC1 seeds sown did not germinate or died before bolting. Figure 4.49 shows the variation in bolting time in the BC1 plants compared to WT, highlighted with an orange dotted line. The WT plants bolted after an average of 99.3 +/-1.04 days, some of the BC lines produced plants that bolted significantly later than WT; especially line BC1_124_3_4, which was comprised of five plants all of which bolted after 115 days or more. Although the BC1 plants were small in number and the potential mutations causing the LB phenotype may be segregating, it appears that a number of the lines are of interest. In an attempt to further remove background EMS mutations, a second BC (BC2) was performed.

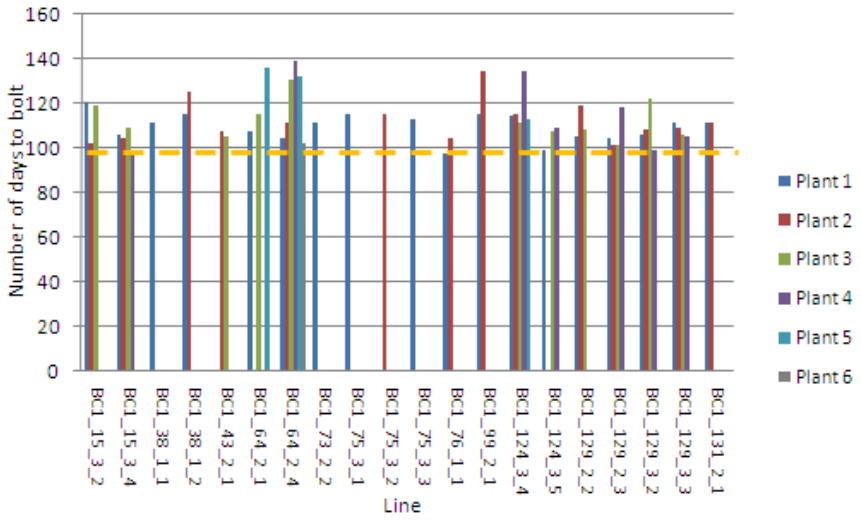


Figure 4.49 – Number of days to bolt for BC1 Saladin plants. Plants representing 11 LB Saladin lines have been BC to WT, BC1 seed was collected and BC1 plants were scored for number of days to bolt. The average number of days to bolt for the WT plants is highlighted with an orange dotted line.

Each of the BC1 plants were BC as many times as possible to WT for a second time. Unlike with the M2 plants used for the first BC to WT, the BC1 plants were far healthier, with better flowers containing more pollen. The amount of BC2 seed produced and collected was much higher than that from the BC1 plants, only line BC1_73_2_2 produced no seed. Therefore BC2 seed was collected from plants representing 10 lines.

To maximise the chances of obtaining homozygous mutant lines, a number of self fertilisations from the BC2 seed collected were set up, the Saladin self fertilisations were performed by Rijk Zwaan®, Fijnaart, Holland, over the winter season under artificial lamps providing a LD photoperiod of 16h. This provided a second growing season per year. The BC2 seed collected from two independent BC2 events from two individual plants from each of the 10 interesting lines was sent to Rijk Zwaan® to be self fertilised, no bolting data was collected from this experiment. The seed collected, BC2_S1, was then screened for homozygous mutant lines.

The bolting time screen of the BC2_S1 plants derived from these 10 LB Saladin lines is discussed in detail in section 4.6.1.10, along with similar data obtained from the 'main' Saladin population.

4.6.1.6 'Main' Saladin EMS population – M1 screen

The 'main' Saladin population was made up of 2612 individual M1 plants. The only difference in the way that the seed was mutagenised to produce the 'test' and 'main' populations was that the seed for the 'main' population was treated with EMS for 18h compared with that of the 'test' population which was treated for 24h.

Each individual M1 seed was sown and transplanted into the field as described in section 2.1.2.1 and 4.2.1, and each plant was scored for days to bolt. Interestingly the amount of phenotypic variation seen between the lines was very low in comparison to the Larissa and Saladin 'test' populations. Only 1012 of the M1 plants grew to bolting, the rest died before bolting; a large proportion of plants were lost because of the wire worm infestation previously mentioned. A spread in bolting time of 55 days was observed, with the earliest plant bolting after 111 days and the latest after 166 days, WT plants bolted after an average of 124.8 +/-0.75 days, see figure 4.50.

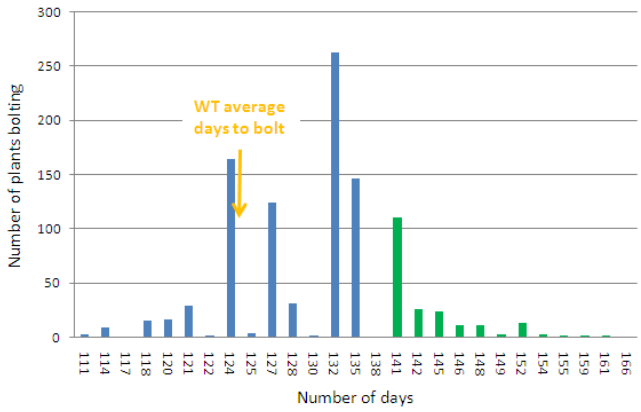


Figure 4.50 –Saladin M1 'main' population number of days to bolt. WT plants bolted after an average of 124.8 +/-0.75 days; orange arrow. Green bars highlight plants bolting significantly later than WT.

To be consistent with the 'test' population the 209 M1 plants (20 % of the 'main' Saladin population), that bolted after 138 days or more; thirteen days later than WT, were classed as being LB, these are highlighted in green in figure 4.50. As these plants are M1 plants, their phenotype could possibly be caused by a dominant mutation. It is important to remember that lines which appear to be LB in the M1 population may be displaying this phenotype because of a mutation carried in the somatic cells which is not inherited through the cells making up the reproductive tissues forming the subsequent M2 population.

Because 209 plants would be difficult to rescreen and BC in the limited space available in the glasshouse, it was decided that only plants which bolted after 148 days or later, nearly two weeks later than WT, would initially be re-screened in the glasshouse. This gave a total of 38 M1 plants, which was a manageable number to screen and BC in the glasshouse, the remainder of the 209 plants would be rescreened as part of the M2 generation screen described in section 4.6.1.8, therefore any lines reproducing their LB phenotype in that screen could be rescreened and BC in the glasshouse at a later date. Of the 38 M1 plants, nine had been severely affected by the wire worm infestation in their early development, the LB phenotype observed in these lines may have been due to this and not because of the affect of the EMS, therefore these lines were not followed up at this stage. Of the remaining 29 LB M1 plants, 26 bolted and set seed. Three lines bolted so late that they did not set seed despite being transplanted and maintained in the glasshouse during the winter. The 26 LB lines bolted between 148 and 161 days are shown in figure 4.51.

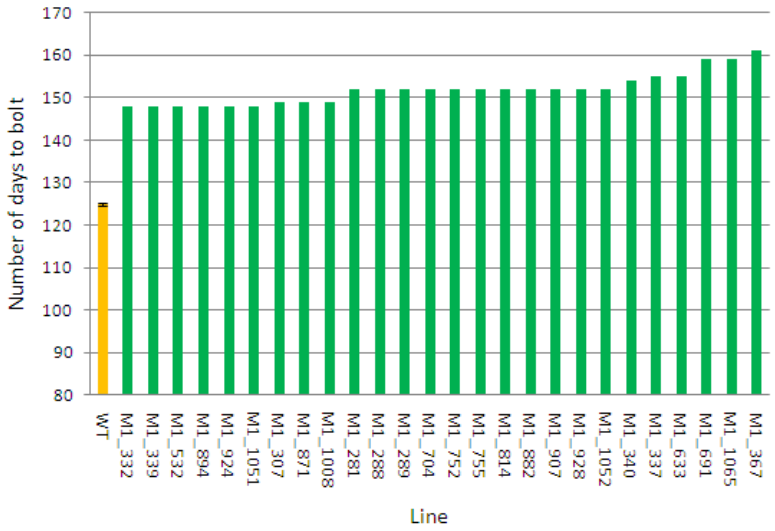


Figure 4.51 – LB M1 plants identified from the 'main' Saladin EMS population.

Four M2 seeds from each of the 26 LB lines identified from the M1 screen were sown and scored in the field, under the protection of haygrove polytunnels under a natural LD photoperiod. WT plants bolted after an average of 125 +/- 0.6 days, an ANOVA test suggested that plants bolting after 138 days or more (as with the M1 screen) were significantly different to WT and were scored as LB (p=<0.001; d.f.=21; l.s.d.=13.4). All four M2 plants grown from seed collected from two of the LB M1 lines; 332 and 871, were LB compared with WT, providing possible evidence of the presence of a dominant mutation causing the LB phenotype, these lines are highlighted in green in figure 4.52. A further five lines; 337, 367, 1051, 1052 and 1065 highlighted in brown in figure 4.52, produced two or three individual plants bolting after or later than 138 days, suggesting the phenotype may also be caused by a dominant mutation. The remaining 19 lines were made up of individual plants which all bolted earlier than the 138 day cut off, the LB phenotype exhibited in these M1 plants may have been caused by an EMS induced mutation in a somatic cell which is not present in the M2 generation.

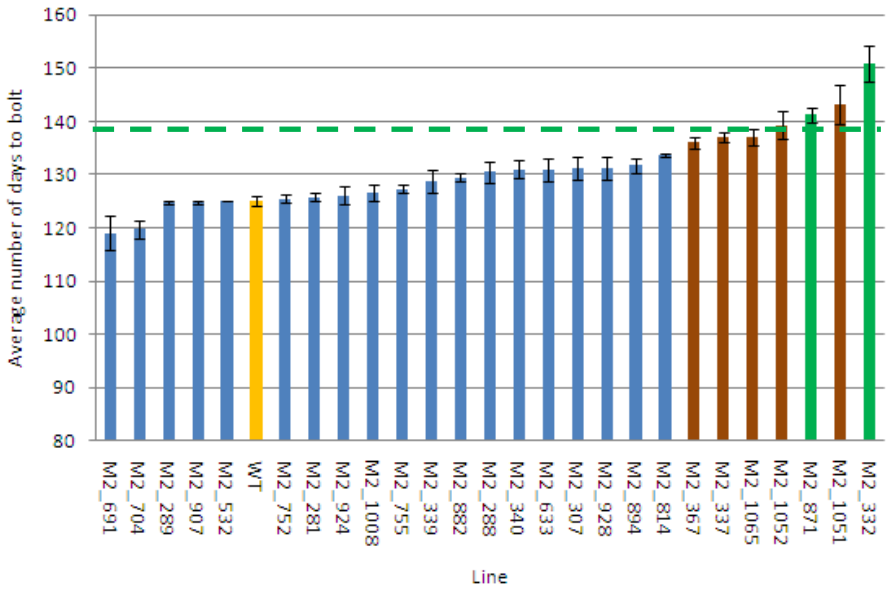


Figure 4.52 – Screen of M2 plants from Saladin 'main' population - based on lines which were LB in M1 generation, the brown bars indicate lines containing two or three LB individuals, the green bars highlight lines were all plants screened were LB compared to WT. Plants bolting after 138 days or more are classed as LB (highlighted with a green dotted line).

4.6.1.7 'Main' Saladin EMS population - Identification of homozygous dominant mutant lines

Four M2 seeds representing five of the seven LB lines were grown for rescoring and possible BC in the glasshouse under a natural LD photoperiod, as described in section 2.1.2.1 and 2.1.2.3. There was limited M2 seed available for lines 332 and 1051, this seed was prioritised for the field trial described in section 4.6.1.8 (and was shown to not have a LB phenotype in that screen), and therefore was not used in this glasshouse experiment. WT plants bolted after an average of 102 +/-2 days, considerably earlier than plants grown in the field, plants bolting after 113.5 days were statistically different from WT ($p=<0.001$; d.f.=11; l.s.d.=11.57). However, as with the previous Saladin glasshouse experiments plants bolting after 112 days or more were scored as LB; this included three plants representing line 367, and single plants from lines 337 and 1065, see figure 4.53. However lines 871 and 1052 produced no LB individuals, this was unexpected as all four M2 plants from line 871

and two plants from line 1052 were LB when scored in the field, see section 4.6.1.6 and figure 4.52.

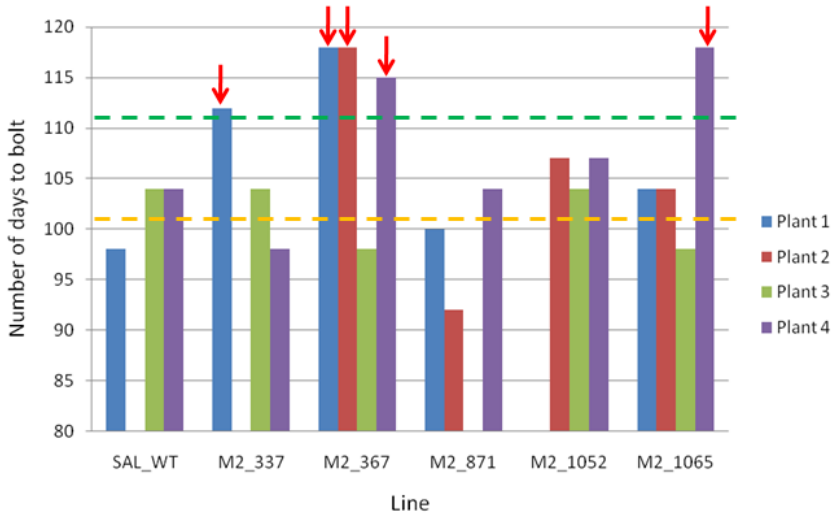


Figure 4.53 – Screen of LB lines in glasshouse. Plants that were LB in the field trial were rescreened in the glasshouse. Five individual plants, highlighted with red arrows, displayed a LB phenotype and were BC to WT. The orange dotted line represents the average number of days that WT plants took to bolt. The green dotted line represents the number of days set for a plant to be LB.

The five individual LB plants; M2_337_1, M2_367_1, M2_367_2, M2_367_4 and M2_1065_4, highlighted with red arrows in figure 4.63, were BC to WT Saladin as described in section 2.2.16. As many BC events as possible were carried out per plant due to the small number of plants available, M2_BC1 seed was then collected. BC, which had been difficult with the 'test' Saladin population was much easier with the plants from the 'main' population; this may be due to the reduced amount of time that the seeds had been treated with EMS.

Where seed was available, four M2_BC1 seeds from four individual BC1 events from the five LB M2 lines were sown and scored for bolting in the glasshouse under a natural LD photoperiod as described in section 2.1.2.1 and 2.1.2.3. Figure 4.54 shows the data collected from the M2_BC1 plants, WT plants bolted after an average of 99.3 +/-1.04 days. A number of individual BC1 plants were significantly LB compared to WT, including all the plants from lines BC1_367_1_4_A and BC1_1065_4_3_A.

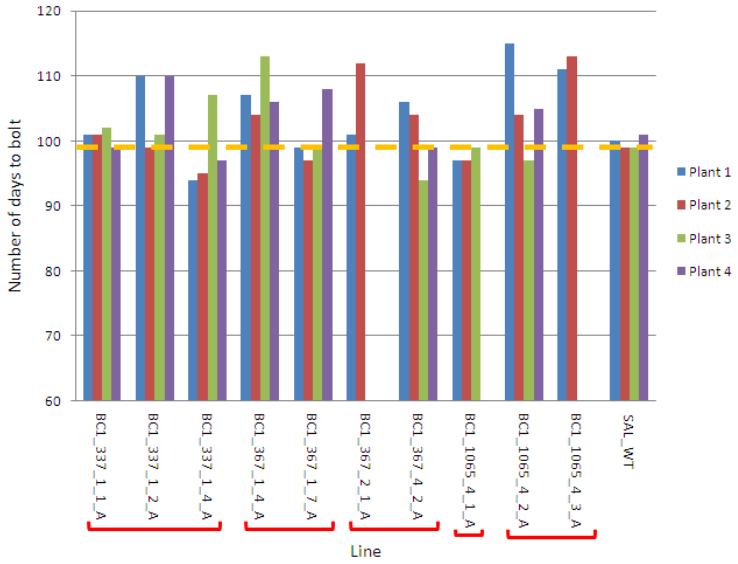


Figure 4.54 – M2 'test' population Saladin plants scored after one round of BC. Some lines are still displaying a LB phenotype compared with the four WT plants. The red 'brackets' highlight the five LB lines that the BC1 seed scored originated from. The orange dotted line corresponds to the average number of days to bolt for WT.

Although these lines are thought to be carrying a dominant mutation causing LB, the plants scored at this stage may not be displaying a LB phenotype because the mutation(s) causing the LB phenotype may have been lost through BC or is segregating within the population. To further remove unwanted background allelic variation caused by the EMS mutagenesis all the plants were BC to WT for a second time. BC2 seed was collected from the plants representing the five LB lines to grow for selfing to produce homozygous mutant lines.

As with the 'test' Saladin population the BC2 seed was grown and self fertilised at Rijk Zwaan®, Fijnaart, Holland over the winter season in the glasshouse under artificial lamps. The seed collected, BC2_S1, was then screened for bolting time, this data is discussed in detail in section 4.6.1.10.

4.6.1.8 'Main' Saladin EMS population – M2 screen

The M2 'main' Saladin population was screened for recessive mutations which would not cause a phenotype when screening the M1 generation, and for any

dominant mutations that were not identified in the M1 generation because of the wire worm infestation. Seed was collected from 868 of the 1012 M1 plants which bolted, described in section 4.6.1.6, the remaining plants died before setting seed, or produced no seed. Four plants from each of the M1 lines (excluding the lines followed up for potential dominant mutations discussed in section 4.6.1.7), that set seed were grown and scored in the field under the protection of haygrove polytunnels as described for the M1 population. A spread in bolting time of 83 days, (73-156 days), was observed, compared to Saladin WT which bolted after an average of 125 +/-0.99 days, (the same number of days as recorded for the M1 population trial). Lines bolting after an average of 130.9 days were significantly different to WT (p=<0.001, d.f.=2401; l.s.d.=5.91). This totaled 249 M2 lines, this amount of lines could not be rescreened and BC in the available space; so a cut off for LB of plants bolting on or after 141 days (around two weeks later than WT, as with previous Saladin field trials) was set. 139 plants (~4 % of the population) were scored as LB; these plants are highlighted in green in figure 4.55.

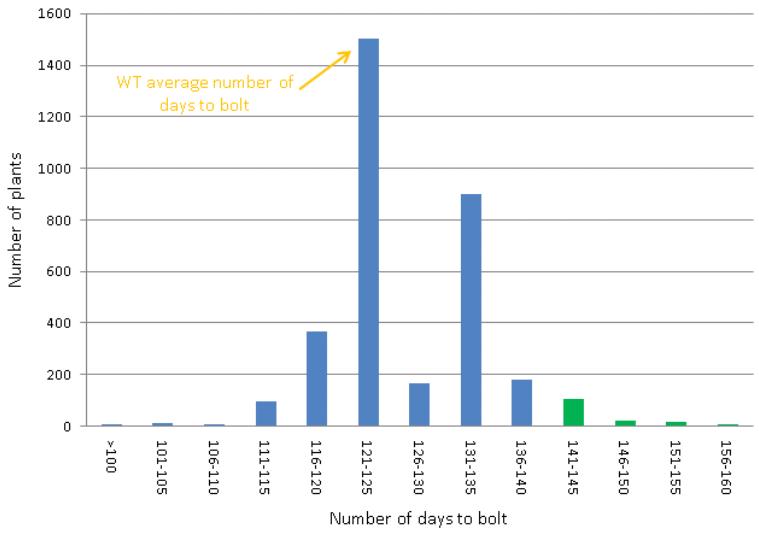


Figure 4.55 – 'Main' Saladin EMS mutagenised M2 population. The graph displays the spread in the number of days that the plants took to bolt.

Q-Q plot analysis of the bolting data was carried out, not all the lines fit the 90 % confidence limits, shown with the red lines in figure 4.56 and highlighted with a red arrow, this means the data does not fit a Gaussian distribution making it difficult to discuss the significance of the LB lines chosen for further analysis.

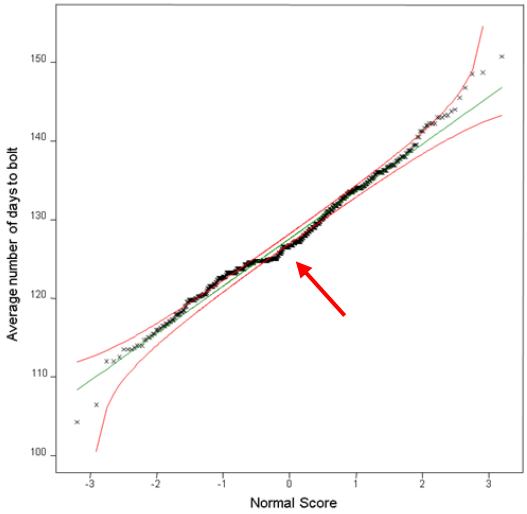


Figure 4.56 – Normal Q-Q probability plot (+ 90 % confidence limits) for Saladin EMS 'main' M2 population. The skewed red lines represent the 90 % confidence level that the data fits a Gaussian distribution, some points (indicated by a red arrow) lie outside this region therefore this population does not fit a Gaussian distribution.

The 139 LB plants represented 81 lines and these were taken forward for further investigation. A breakdown of these plants can be seen in table 4.6, five lines looked particularly interesting, all four plants grown representing these lines bolted later than WT. 50 lines contain one to three individual plants bolting later than WT. A further 26 individual plants representing LB lines did not produce any seed. This left a total of 55 lines to follow up.

	Number of lines
LB individuals, but no seed set	26
1 LB individual	29
2 LB individuals	16
3 LB individuals	5
All 4 plants LB	5
Total	81

Table 4.6 – LB Saladin lines in the M2 population. The 81 lines represented contained one-four individual LB plants.

4.6.1.9 ‘Main’ Saladin EMS population - M3 Screen and back-crossing

M3 seed was collected from the M2 plants of the 55 LB lines from the ‘main’ Saladin EMS population scored in the field. Of the 55 M2 lines, 29 lines contained just one plant which was LB compared to WT; four seeds from each LB plant was scored for bolting. 16 lines contained two plants which were LB compared to WT, again four seeds from each LB plant were scored. Three seeds per LB plant from the five lines containing three LB plants were also scored, along with two seeds from each of the four plants from the five lines which were all LB. All of these plants were grown in the glasshouse, under a natural LD photoperiod.

The overall data collected from the M3 generation containing lines with one, two and three LB individuals will be discussed initially, as screening of these lines took place at a different time from the lines containing all four plants with a LB phenotype. Of the plants scored, only line 364 did not produce plants which bolted. WT plants bolted after an average of 81 +/-0.48 days, ANOVA suggested that plants bolting after 89 days are significantly different to WT ($p=<0.001$; d.f.=131; l.s.d.=8.284), a difference of only eight days later than WT as opposed to ten days as in previous Saladin glasshouse experiments. As this screen took place in the middle of summer, all plants bolted earlier than Saladin plants in other glasshouse trials – WT plants after an average of 81 +/-0.48 days compared to 102 +/-2 days, therefore the ANOVA level of significance was accepted.

There were 29 M2 lines of which one of the four plants sown was LB compared with WT, of the four M3 seeds sown for each of the 29 lines, seven produced plants which were all LB, a further 16 produced one to three LB individuals five lines contained no LB individuals, plants representing the final line all died, see figure 4.57.

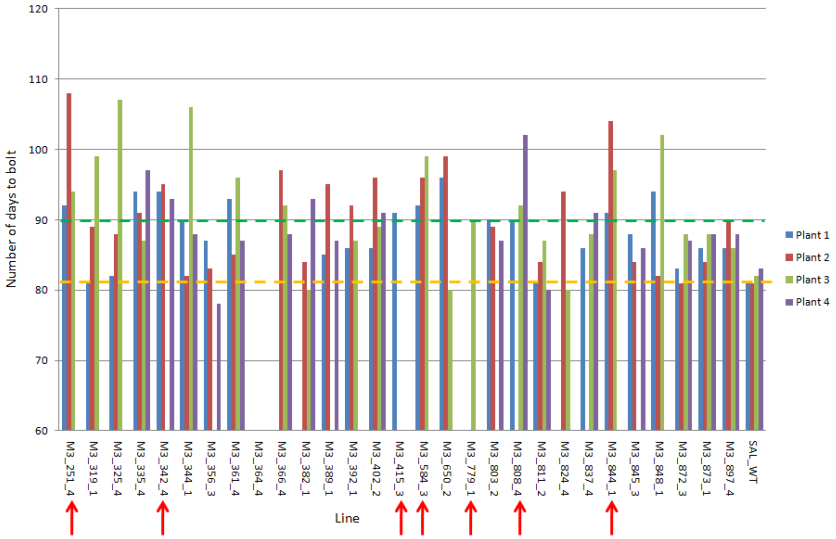


Figure 4.57 – Breakdown of bolting time for each M3 line comprised of seed collected from M2 lines containing only one individual LB plant. The average number of days to bolt for WT is highlighted with an orange dotted line, the number of days for a plant to be classified as LB is highlighted with a green dotted line. The red arrows indicate lines where all plants screened were LB.

16 M2 lines contained two LB individual plants of the four that were grown, M3 seed was collected from both LB individuals for 11 of the lines, and from only one plant for the other five (the remaining plants died or produced no seed), making a total of 27 plants. Four M3 seed from each of the 27 plants were sown, the progeny from eight of the M2 plants were all LB, 12 produced one to three LB individuals and seven contained no LB individuals, highlighted with red arrows in figure 4.58.

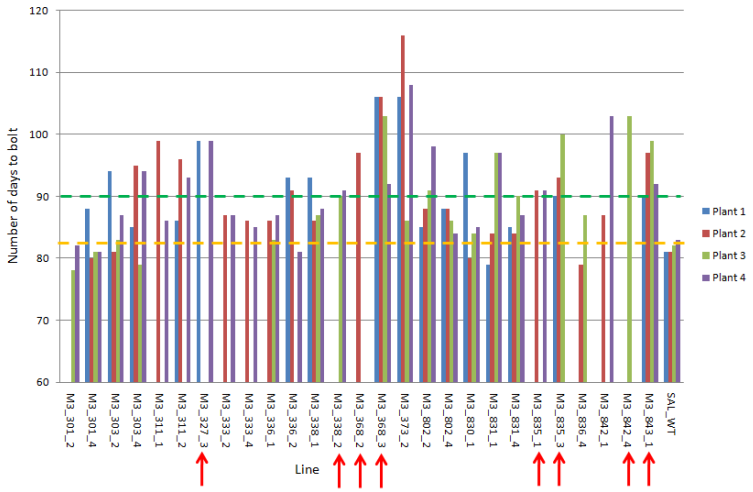


Figure 4.58 - Breakdown of bolting time for each M3 line comprised of seed collected from M2 lines containing two individual LB plants. The average number of days to bolt for WT is highlighted with an orange dotted line, the number of days for a plant to be classified as LB is highlighted with a green dotted line. The red arrows indicate lines where all plants screened were LB.

Finally, there were five M2 lines containing three LB individuals, three M3 seed from each of the LB individuals were sown and scored for bolting. Five of the M2 lines had M3 progeny that were all LB, a further four M2 lines produced one to three LB M3 individuals, five M2 lines produced no LB M3 individuals, and the plants derived from the final line (M3_372_3) all died before bolting, see figure 4.59.

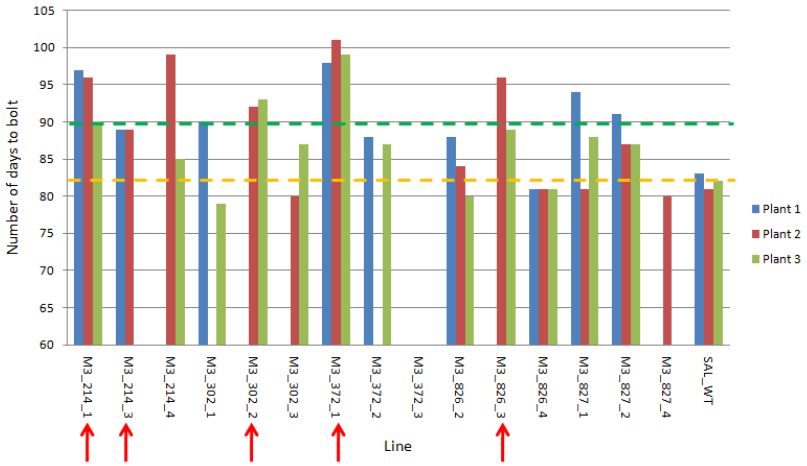


Figure 4.59 - Breakdown of bolting time for each M3 line comprised of seed collected from M2 lines containing three individual LB plants. The average number of days to bolt for WT is highlighted with an orange dotted line, the number of days for a plant to be classified as LB is highlighted with a green dotted line. The red arrows indicate lines where all plants screened were LB.

There were also five M2 lines which produced four LB individual plants, two M3 seeds from each of the four plants were sown and scored. Figure 4.60 shows the data observed from these lines; WT plants bolted after an average of 99.3 +/-1.04 days, plants which bolted after 110 days, ten days later than the last WT plant were scored as LB, no ANOVA test was performed on this data as there are only two replicates of each. Five of the six plants scored from line 353 were LB, see figure 4.60. However the other lines did not reproduce the phenotype observed in the M2 field trial. Only one other M2 plant, from line 829 produced progeny which all contained LB plants. M2 seed collected from five LB M2 plants produced plants which died before bolting, see figure 4.60 for a breakdown of the bolting time for the individual plants.

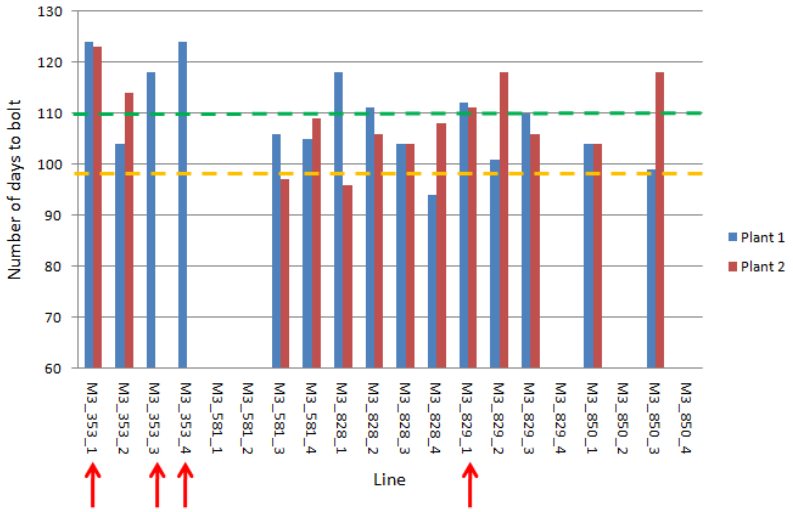


Figure 4.60 – Breakdown of bolting time for each M3 line comprised of seed collected from M2 lines containing four individual LB plants. The average number of days to bolt for WT is highlighted with an orange dotted line, the number of days for a plant to be classified as LB is highlighted with a green dotted line. Five of the six plants which bolted representing line 353 bolted significantly later than WT. The red arrows indicate lines where all plants screened were LB.

In summary, the 'main' Saladin M2 screen produced a number of interesting lines with one to four of the plants representing each line displaying a LB phenotype. The subsequent screening of the resulting M3 plants indicated that 46 out of 55 independent lines, originating from individual M1 plants, contained plants with LB phenotypes (24 lines originating from seed collected from independent M2 plants produced M3 plants which all had a LB phenotype, indicated with red arrows in figures 4.67-4.70). All M3 plants with a LB phenotype were BC to WT, as many independent BC events per plant were carried out as possible.

12 of the 24 lines where all the M3 plants were LB, (lines 251, 335, 338, 342, 368, 372, 584, 808, 835, 842, 843 and 844), were prioritised for BC to WT. The data collected for each of these lines from each generation gives a strong indication of the possible genetics underlying the bolting phenotype recorded.

Of these 12 lines, 5 lines 251, 342, 368, 584 and 835 all appear to be caused by a dominant mutation because of the segregation of the LB phenotype in the M1, M2 and M3 generations. Each of the M1 plants of these lines were LB compared to WT,

(bolting after 135, 146, 135, 132 and 146 days respectively, compared with WT which bolted after an average of 124.8 +/-0.75 days), although they did not bolt after the cut off point of 148 days which was used to identify LB lines in the M1 screen (section 4.6.1.6). In the M2 generation, only four plants were screened per line but a rough 3:1 ratio for LB plants to plants bolting as WT was observed, the M3 plants grown from seed collected from the LB M2 plants were all LB. (A further three lines; lines 327, 415 and 779 also appeared to be LB due to a dominant mutation; however BC was unsuccessful for these lines). The LB phenotype of a further two lines; lines 338 and 842 containing plants all bolting later than WT in the M1 and M3 generations, although not all bolting as late as the 98 day cut off for LB set for the M3 generation, may also be caused by a dominant mutation.

Of the remaining lines; lines 372, 808, 843 and 844 the LB phenotype is potentially caused by a recessive mutation. The M1 plants screened for each line bolted as WT. The four M2 plants scored for each line produced 3:1, 1:3, 2:2 and 1:3 ratios respectively for LB plants to plants bolting as WT. A 1:3 ratio would be expected here, but may not been seen due to the small number of plants grown. In the M3 generation derived from LB M2 plants (which would have to be homozygous for the recessive mutation to have a LB phenotype) all plants scored for lines 372, 808, 843 and 844 were LB compared to WT, as expected if they were derived from homozygous mutant plants. A further line; line 335 bolted as WT in the M1 generation and produced a 1:3 LB plants to plants bolting as WT ratio in the M2 generation, the four M3 plants derived from the LB M2 plant were also all later bolting than WT, however one failed to make the 98 day cut off. These five 'recessive mutant' lines were also prioritised. In total 12 LB Saladin lines identified from the M2 'main' population screen appear to be good lines to follow up as the

mutation causing the LB phenotype appear to be segregating as a single allele within the population. The remaining LB lines are more complicated as they are not exhibiting a single gene segregation pattern, this may be because there is more than one mutation that affects bolting time that is segregating in these lines.

Seven lines; lines 251, 338, 342, 368, 584, 835 and 842 whose LB phenotype was possibly caused by a dominant mutation, and which set BC1 seed, along with five lines; 335, 372, 808, 843 and 844 whose LB phenotype was possibly caused by a recessive mutation were prioritised for selfing. BC1 seed was collected from these plants and was sent for selfing over the winter season, under artificial lamps at Rijk Zwaan®, Fijnaart, Holland.

The BC1_S1 seed was screened and is discussed along with the data obtained from the BC2_S1 seed from the 'main' and 'test' populations in section 4.6.1.10.

4.6.1.10 Screening the Saladin BC and selfed lines for homozygous LB lines

The 25 interesting LB Saladin lines that have been BC once or twice to WT were selfed at Rijk Zwaan®. These comprise 10 LB lines (lines 15, 38, 43, 64, 75, 76, 99, 124, 129 and 131), obtained from the 'test' Saladin EMS population, (see section 4.6.1.4 and figure 4.59 for details), and 3 LB lines (lines 337, 367 and 1065), from the 'main' Saladin EMS population (see section 4.6.1.7 and figure 4.64 for details) which have been BC twice before selfing. The remaining 12 LB lines (251, 335, 338, 342, 368, 372, 584, 808, 835, 842, 843 and 844) from the 'main' Saladin EMS population were BC once to WT before selfing, see section 4.6.1.9.

Six plants representing the selfed seed collected from two independent BC events from two individual BC1/BC2 plants (where seed was available) were grown in the glasshouse under a natural LD photoperiod and were scored for days to bolt.

Figure 4.61 show a breakdown of the data obtained for the BC2_S1 plants obtained from the 'test' Saladin EMS population. WT plants bolted after an average of 82.3 +/-1.78 days, plants bolting ten days later than WT, after 92 or more days were scored as LB. Because the mutations affecting bolting time within these populations are still segregating no statistical analysis has been performed on this data.

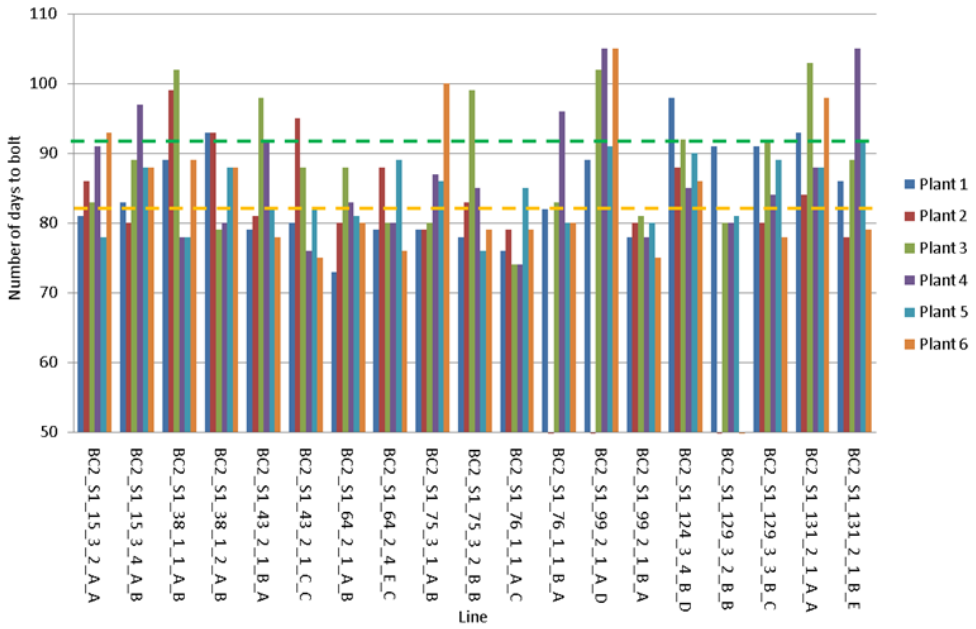


Figure 4.61 – BC2_S1 plants from the 'test' Saladin population. The orange dotted line indicates the average number of days to bolt for WT and the green line indicates the cut off for LB lines.

23 individual plants bolted after 92 or more days, interestingly no seed from either of the two independent BC2 events from line 64 produced LB BC2_S1 plants, lines 76_1_1_A_C, 99_2_2_1_B_A and 129_3_2_B_B also produced no LB individuals. The lack of LB plants observed in these lines is unexpected as the mutation causing LB is believed to be dominant in these lines, see section 4.6.1.5 suggesting the mutation affecting the phenotype must have been lost through BC.

Figure 4.62 displays the data collected from the BC2_S1 plants obtained from the 'main' Saladin EMS population.

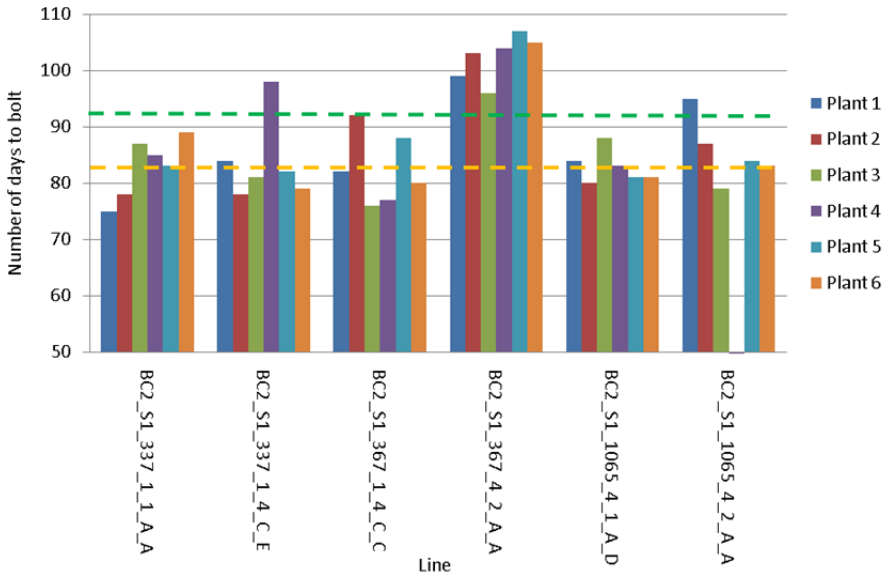


Figure 4.62 – BC2_S1 plants from the 'main' Saladin population. The orange dotted line indicates the average number of days to bolt for WT and the green line indicates the cut off for LB lines.

Nine plants bolted after 92 days or more and are therefore scored as LB. Of particular interest are the six plants making up line 367_4_2_A_A which all bolt later than WT. No plants representing lines 337_1_1_A_A and 1065_4_1_A_D are LB; the LB phenotype of these lines is believed to be caused by a dominant mutation see section 4.6.1.7, and so it is surprising that there are no LB individuals recorded for these lines.

Figure 4.63 show the data collected from the BC1_S1 plants from the 'main' Saladin EMS population.

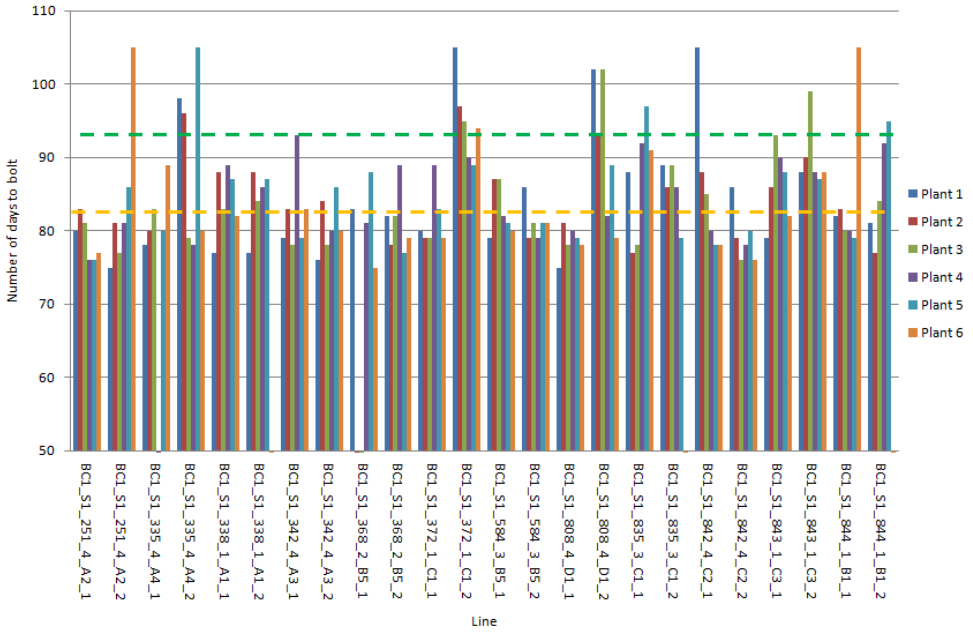


Figure 4.63 – BC1_S1 plants from the 'main' Saladin population. The orange dotted line indicates the average number of days to bolt for WT and the green line indicates the cut off for LB lines.

20 plants bolted after 92 days or more and were scored as LB. Lines 338, 368 and 584 all produced no LB lines, suggesting the mutation (likely to be a dominant, see section 4.6.1.9) causing the LB phenotype in each line could have been lost through BC. Lines 251_4_A2_1, 335_4_A4_1, 372_1_C1_1, 808_4_D1_1, 825_3_C1_2 and 842_4_C2_2 also contained no LB individuals. It has been discussed in section 4.6.1.9 that the mutations causing the LB phenotype in lines 335, 372 and 808 are likely to be recessive, therefore it is not surprising that no LB individuals were observed in the lines as the mutation may still be segregating.

In total 32 LB BC2_S1 plants from the 'test' and 'main' populations and 20 LB BC1_S1 plants from the 'main' population have been identified. Seed will be collected from each of these plants and it will be screened to identify homozygous mutant lines. These lines can then be tested with a vernalisation treatment to see if the bolting phenotype can be recovered, which would suggest a mutation within an autonomous pathway gene. They can also be grown in the field to verify the LB phenotype under commercial growing conditions.

4.7 Retrotransposon Lines

4.7.1 Introduction

A number of lines of a cultivated butterhead lettuce; *L. sativa* cv. Jessy which had been transformed with the tobacco retrotransposon element, *Tnt1,* by Dr Mazier's group, Centre de Recherche Agronomique d'Avignon, France (Mazier *et al*., 2007) were made available to screen for bolting time. These lines had multiple *Tnt1* elements randomly inserted within the lettuce genome, making them a useful resource for gene tagging experiments. The rationale behind screening this population was that any variation in bolting time identified could be followed up by identifying the genes into which the retrotransposon element had inserted itself. Any gene identified, known to be involved in flowering through research in *Arabidopsis,* could be investigated in more detail.

4.7.2 Materials and Methods

Jessy seeds collected from plants that had been transformed with the tobacco retrotransposon element *Tnt1* by Dr Mazier's group were made available to us for screening for bolting time. Section 1.6.3.2, contains a breakdown of the seeds which were screened. Dr Mazier's group transformed seed (T0), which was sown, T1 seed was collected from these plants. T1 seed was then sown and T2 seed was collected from these plants.

The Jessy seed made available to us was comprised nine lines, including WT. Seed from only one T1 Jessy plant was made available; J_Tnt8a, which was screened along with the T2 seed collected by Dr Mazier's group from two J_Tnt8a T1 plants; plants 1 (p1) and 3 (p3). Seed from a further five T2 plants collected by Dr Mazier's

group, originating from two independent T1 lines was also scored for days to bolt, see table 4.7 for details.

Line	T1 lines	T2 line
Jessy	Tnt8a	Tnt8a_p1
		Tnt8a_p3
	N/A	Tnt1a_p2
		Tnt1a_p5
		Tnt1a_p6
	N/A	Tnt6a_p6
		Tnt6a_p8

Table 4.7 – Details of the *L. sativa* lines transformed with the Tobacco *Tnt1* retrotransposon element screened for bolting time. N/A means seed not available.

Six seeds from each Jessy line, including WT were sown, as described in section 2.1.2.1, in duplicate and were transplanted into the glasshouse as described in section 2.1.2.3. The first duplicate was given a vernalisation treatment of four weeks at 4 °C. The second duplicate did not have a vernalisation treatment. All of the plants were grown under an artificial LD photoperiod of 16 h and were scored for days to bolt, T2 or T3 seed from each line was collected. This experiment was repeated under a natural LD photoperiod, this time, eight plants from each line had a vernalisation treatment and eight had no treatment. Lines which showed an interesting LB phenotype were BC to WT in an attempt to reduce the number of retrotransposon elements within the line, while keeping the element affecting bolting time.

Four seed (BC1), collected from two independent BC events from two individual plants per LB line were sown, with no vernalisation treatment. The plants were grown under an artificial LD photoperiod in the glasshouse and were scored for days to bolt. Self seed (BC1_S1), from each plant was collected. In an attempt to identify homozygous LB mutant lines, four BC1_S1 seed from two selfed plants, originating from independent BC1 events representing the LB lines were grown under a natural LD photoperiod in the glasshouse. This time two set of plants were sown, one set

was vernalised and one set were untreated and the plants were scored for days to bolt. The strategy to obtain homozygous LB mutant Jessy lines is illustrated in figure 4.64.

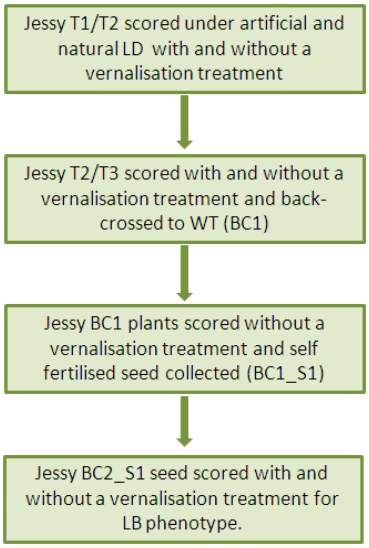


Figure 4.64 - Strategy to obtain homozygous LB Jessy lines

4.7.3 Results

4.7.3.1 Jessy population – initial screens

The overall germination rate for the seeds provided by Dr. Mazier was poor, so the initial data set was limited.

WT plants bolted after an average of 56.3 +/-0.8 days with no vernalisation treatment. All lines screened with the exception of line J_T2_Tnt8a_p3 bolted significantly later than WT when non-vernalised ($p=<0.001$, d.f.=15, l.s.d.=4.52). However, lines bolting on average seven days or later than WT were scored as LB for the purposes of this project. This meant that five lines were LB compared to WT under non-vernalised conditions. Four of the five LB lines; lines J_T2_Tnt8a_1, J_T2_Tnt1a_2, J_T2_Tnt1a_5 and J_T2_Tnt1a_6 (highlighted with red arrows in figure 4.65) were on average 7.67, 7.67, 14.34 and 19.7 days later to bolt than WT with no vernalisation treatment respectively. However when these lines were

subjected to a vernalisation treatment they were no longer significantly LB compared to WT (p=<0.001, d.f.=28, l.s.d.=3.15), their bolting phenotype had been restored to WT. WT plants bolted, on average after 45.7 +/-1.65 days with a vernalisation treatment; lines J_T2_Tnt8a_1, J_T2_Tnt1a_2, J_T2_Tnt1a_5 and J_T2_Tnt1a_6 bolted after an average of 45 +/-4, 46 +/-1.39, 48.8 +/-0 and 46 +/-2.33 days respectively. This was of interest as genes involved in the autonomous pathway, which are being targeted in this project, when mutated in *Arabidopsis* result in a LB phenotype which can be restored to a WT bolting phenotype with a vernalisation treatment. T1 seed representing the T2 J_T2_Tnt1a lines was not made available to screen in this project. The other LB line; line J_T1_Tnt8a, (highlighted with a green arrow in figure 4.70), was represented by only one germinated plant under both non-vernalised and vernalised conditions. Regardless of a vernalisation treatment each plant bolted significantly later than WT, (eight days later when non-vernalised and 12.3 days later when vernalised). This line was also represented in the screen by two T2 lines; J_T2_Tnt8a_1 and J_T2_Tnt8a_3, progeny collected by Dr Mazier's group from J_T1_Tnt8a in a previous screen. Neither of these lines reproduced the LB phenotype recorded with the T1 line; line J_T2_Tnt8a_3 bolted on average at a similar time to WT whether subjected to a vernalisation treatment or not.

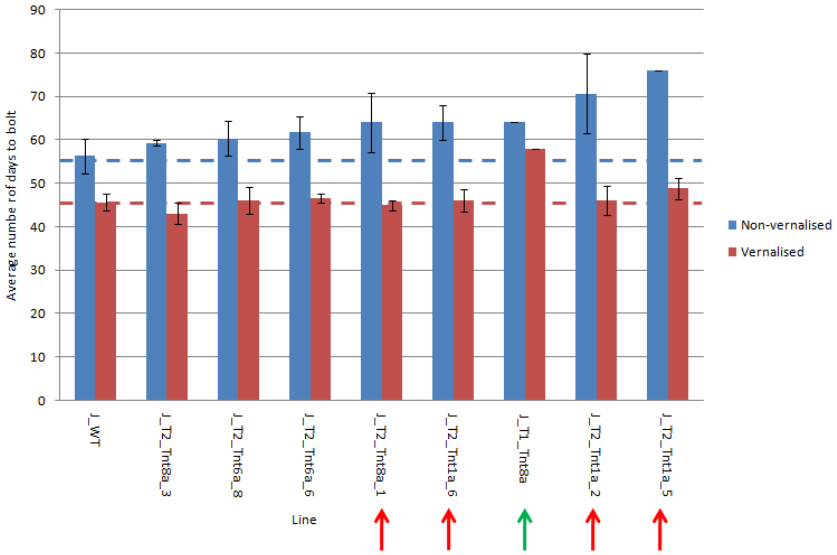


Figure 4.65 – Average number of days to bolt for Jessy lines containing *Tnt1* elements scored under an artificial LD photoperiod. Plants were grown with and without a vernalisation treatment. The red arrows indicate lines of interest with a LB phenotype when not subjected to a vernalisation treatment, this phenotype is restored to WT when a vernalisation treatment is applied. The green arrow indicates a line which is LB regardless of a vernalisation treatment.

This experiment was repeated under a natural LD photoperiod to see if any of the interesting lines reproduced the bolting phenotype recorded under an artificial LD photoperiod. Because the germination rate was poor for the seeds planted in the first screen, self seed from the latest bolting plants scored from each line in the first screen was collected and used in the repeat experiment. 16 plants were scored in this screen, eight of which were subjected to a vernalisation treatment, while eight had no treatment. All plants bolted earlier under natural light LD conditions than under artificial light conditions as observed in the Larissa experiments. WT plants bolted after an average of 48.5 +/-0.33 days when not vernalised. The only line which was significantly LB compared to WT when non-vernalised was J_T3_Tnt1a_5 ($p=<0.001$, d.f.=34, l.s.d.=2.98), it was also the only line that bolted on average seven days or more later than WT. Plants representing the two other LB lines which originated from the same T0 plant, J_T3_Tnt1a_2 and J_T3_Tnt1_6, bolted on average 3.64 and 3.8 days later than WT when non-vernalised. Line J_T3_Tnt8a_1,

which was LB in the first screen, bolted just 1.8 days later on average than WT when non-vernalised.

The LB phenotype of line J_T3_Tnt1a_5 was restored to a WT bolting phenotype when the T3 plants were subjected to a vernalisation treatment, consistent with the results from the first screen under an artificial LD photoperiod. Interestingly lines J_T3_Tnt1a_2 and J_T3_Tnt1a_6 which also bolted later than WT (although less than seven days later than WT) when non-vernalised, also bolted as WT when vernalised. These three lines are highlighted with red arrows in figure 4.66, each of the three lines showed no significant difference in days to bolt when compared to WT when vernalised (p=0.116, d.f.=31, l.s.d.=1.24).

T2 seed collected from line J_T1_Tnt8a which was LB with and without a vernalisation treatment in the initial screen, did not produce any LB plants under a natural LD photoperiod. The LB phenotype present in the T1 plants observed in the first screen appears to have been lost in the T2 generation, this is interesting as none the T3 seed collected from the LB J_T2_Tnt8a_1, which was LB in the first screen, produced plants which lost the LB phenotype under a natural LD photoperiod.

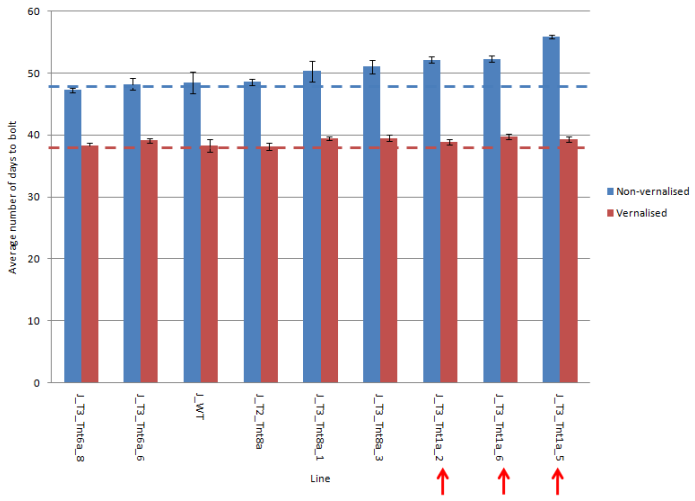


Figure 4.66 - Average number of days to bolt for Jessy lines containing *Tnt1* elements scored under a natural LD photoperiod. Plants were grown with and without a vernalisation treatment. The red arrows confirm three interesting lines with a LB phenotype when not subjected to a vernalisation treatment, a phenotype rescued when a vernalisation treatment is applied.

4.8.3.2 Jessy population – Identification of homozygous mutant lines

The plants from each of the three confirmed LB Jessy lines were BC to Jessy WT in an attempt to remove some of the background retrotransposon elements in the Jessy lines which were not causing the LB phenotype. The four non-vernalised plants from each line which showed the greatest delay in bolting were BC to WT Jessy as many times as possible. BC1 seed was collected from these plants; line J_T2_Tnt1a_6 did not produce any BC1 seed and therefore could not be followed up. Four BC1 seed, collected from two independent BC events from two individual plants per LB line were sown and the plants were scored for days to bolt under an artificial LD photoperiod in the glasshouse. The parental LB lines J_T3_Tnt1a_2 and _5 were also sown as controls together with WT.

Approximately half the BC1 plants representing each of the two lines J_T3_Tnt1a_2 plants 1 and 3 bolted around the same time as the average number of days that WT took to bolt; 71.25 days. The other half bolted seven days or more later than the average number of days to bolt for WT. Overall the plants screened bolted between 63 and 85 days, see figure 4.67. All control J_T3_Tnt1a_2 plants bolted later than the average number of days recorded for WT. The 1:1 segregation of LB to plants bolting as WT in the BC of line J_T3_Tnt1a_2 suggests the presence of a heterozygous dominant mutation caused by the retrotransposon element in this line.

All but one of the BC1 plants from the two independent BC events collected from two LB J_T3_Tnt1a_5, plants 2 and 6 bolted later than 78 days. The BC1 plants screened originating from the J_T3_Tnt1a_5 plants bolted between 75-96 days. This strongly suggests that the mutation causing the LB phenotype observed in J_T3_Tnt1a_5 is a dominant mutation and is homozygous in this line. All plants screened from this line were LB compared to WT in the T2 and T3 generations, as

well as after a single round of BC. Self seed was collected from the BC1 plants (BC1_S1). If the mutation affecting the LB phenotype is dominant then it would be expected that a 3:1 ratio of LB to plants bolting as WT would be seen in the BC1_S1 generation. The J_T3_Tnt1a_5 control plants were all LB compared with the WT plants screened.

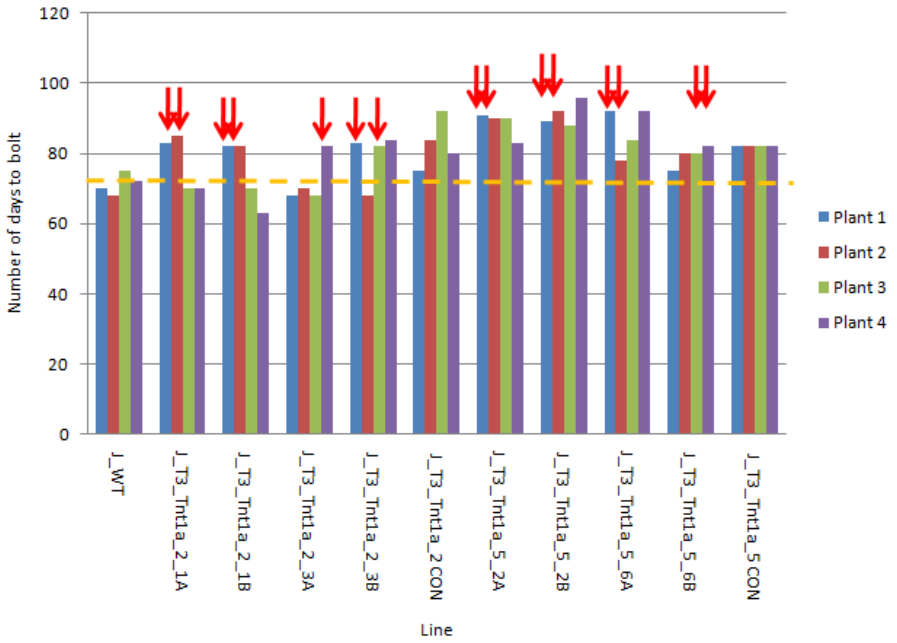


Figure 4.67 - Number of days to bolt for Jessy lines which have been BC to Jessy WT, screened under an artificial LD photoperiod. LB Jessy lines; J_T3_Tnt1a _2 plant 1 and 3 and J_T3_Tnt1a _5 plant 2 and 6 were BC to Jessy WT, four seed from two independent BC events were screened for number of days to bolt. The orange dotted line indicates the average number of days to bolt for WT plants. The red arrows indicate the plants from which BC1_S1 seed was screened.

Because the two LB Jessy lines both originated from the same T0 plant it would be expected that they would contain the same retrotransposon element insertions at the same gene locations. From the BC1 data it can be inferred that the same mutation is causing the LB phenotype; the mutation is homozygous in the T3_Tnt1a_5 plants, but heterozygous in T3_Tnt1a_2 plants.

Four BC1_S1 seed were collected from two LB BC1 plants (where two LB individuals were identified) and these were grown to be screened for bolting time. The selected BC1 plants from which the BC1_S1 seed was collected are highlighted

with red arrows in figure 4.72; these were J_T3_Tnt1a_2_1A plants 1 and 2, J_T3_Tnt1a_2_1A plants 1 and 2, J_T3_Tnt1a_2_3A plant 4 and J_T3_Tnt1a_2_3B plants 1 and 3, as well as J_T3_Tnt1a_5_2A plants 1 and 2, J_T3_Tnt1a_5_1B plants 1 and 2, J_T3_Tnt1a_5_6A plants 1 and 2 and J_T3_Tnt1a_5_6B plants 3 and 4.

All plants were grown under a natural LD photoperiod in the glasshouse. Two sets of plants were sown, one set was vernalised. The plants were scored for number of days to bolt, non-vernalised WT plants bolted after an average of 46.25 +/-0.25 days, BC1_S1 plants bolting on or after 53 days (seven days later than WT), were scored as LB. Figure 4.68 shows the data collected from the non-vernalised plants. For each of the BC1_S1 lines grown an approximate 3:1 ratio of LB plants to plants bolting as WT is observed (approximate because only four plants were grown), table 4.8 summarises each line in detail.

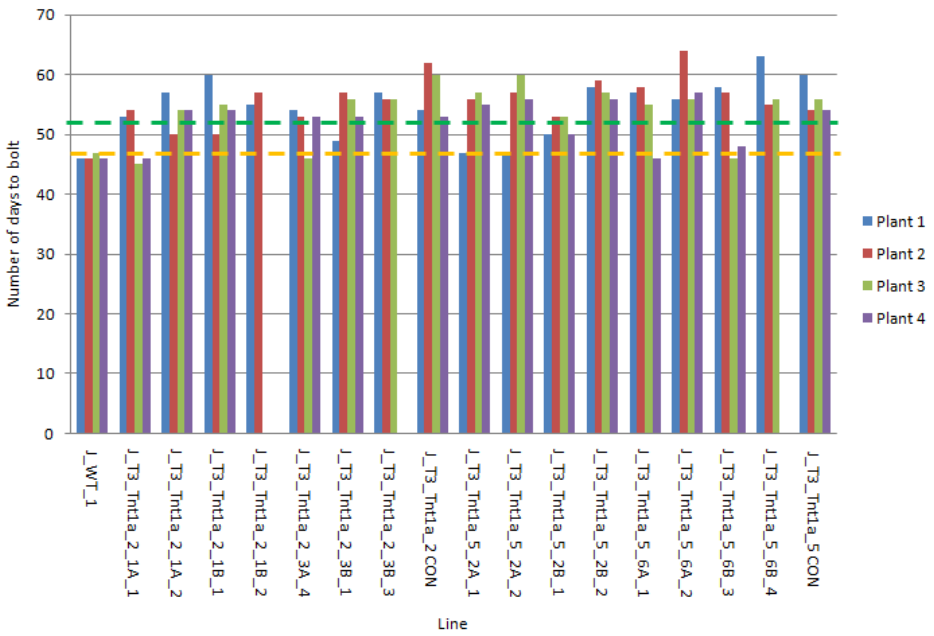


Figure 4.68 - Number of days to bolt for LB Jessy lines which have been BC to Jessy WT and selfed, screened under a natural LD photoperiod with no vernalisation treatment. Four seed from two independent BC1 events per plant were screened for number of days to bolt (except line J_T3_Tnt1a_2_3A where only one BC1 plant was LB). The orange dotted line indicates the average number of days to bolt for WT plants. The green dotted line indicts the cut off point for plants to be classified as LB.

Line	LB:WT plants	Line	LB:WT plants
J_T3_Tnt1a_2_1A_1	1:1	J_T3_Tnt1a_5_2A_1	3:1
J_T3_Tnt1a_2_1A_2	3:1	J_T3_Tnt1a_5_2A_2	3:1
J_T3_Tnt1a_2_1B_1	3:1	J_T3_Tnt1a_5_2B_1	1:1
J_T3_Tnt1a_2_1B_2	1:1	J_T3_Tnt1a_5_2B_2	All LB
J_T3_Tnt1a_2_3A_4	3:1	J_T3_Tnt1a_5_6A_1	3:1
J_T3_Tnt1a_2_3B_1	3:1	J_T3_Tnt1a_5_6A_2	All LB
J_T3_Tnt1a_2_3B_3	All LB	J_T3_Tnt1a_5_6B_3	1:1
J_T3_Tnt1a_2 CON	All LB	J_T3_Tnt1a_5_6B_4	All LB
		J_T3_Tnt1a_5_CON	All LB

Table 4.8 – Summary of ratios observed for LB plants to plants bolting as WT for Jessy lines J_T3_Tnt1a_2 and J_T3_Tnt1a_5 after one round of BC and a self fertilisation under non- vernalised conditions.

The BC1_S1 vernalised plants all bolted between 32 and 44 days, at a similar time to the average number of days recorded for WT of 37.25 +/-1.18 days, see figure 4.69 suggesting the vernalisation treatment has restored the bolting phenotype to WT. One-way ANOVA analysis backs this up, only line J_T3_Tnt1a_5_6A_2 showed any significant difference to the number of days to bolt compared with WT (p=0.079, d.f.=50, l.s.d.=3.28), the line contains two individual plants which bolted seven days later than WT, a point at which they would be scored as LB.

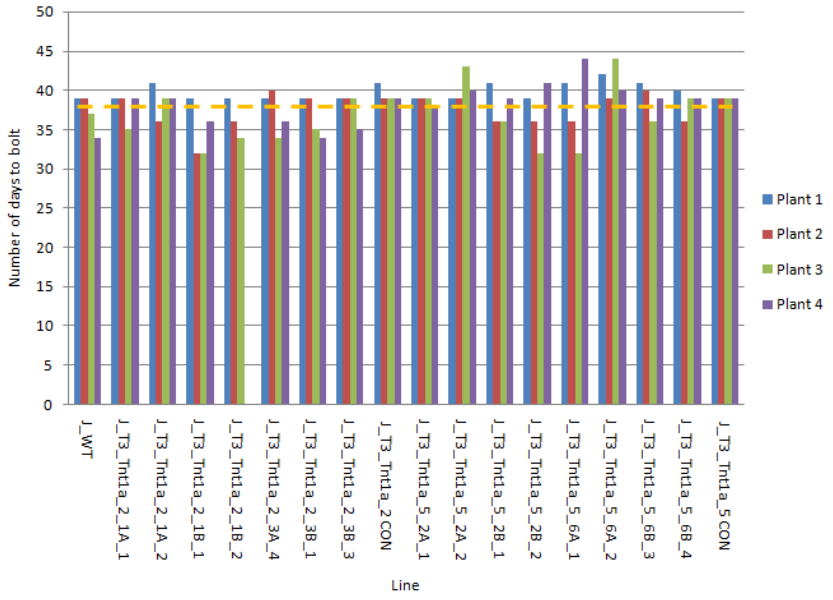


Figure 4.69 - Number of days to bolt for LB Jessy lines which have been BC to Jessy WT and selfed, screened under a natural LD photoperiod with a vernalisation treatment. Four seed from two independent BC1 events per plant were screened for number of days to bolt (except line J_T3_Tnt1a_2_3A where only one BC1 plant was LB). The orange dotted line indicates the average number of days to bolt for WT plants.

Figure 4.70 shows BC1_S1 plants from line J_T2_Tnt1a_5_6A_1; under non-vernalised conditions the plant is LB compared to the WT plant in the photo, however a plant from the same line bolts at the same time as WT when subjected to a vernalisation treatment.

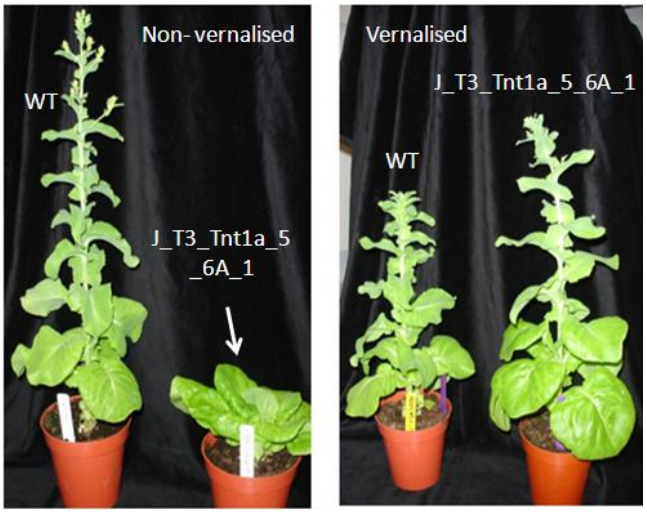


Figure 4.70 – Example of LB Jessy line J_T3_Tnt1a_5_6 after one round of BC to WT and a self fertilisation. The photo on the left shows plants which have not been subjected to a vernalisation treatment; the WT plant bolted after 46 days and the J_T3_Tnt1a_5_6A plant bolted after 63 days. The photo on the right shows plants grown with a vernalisation treatment; the bolting phenotype of the J_T3_Tnt1a_5_6A plant is restored to WT, both plants bolted after 39 days.

The segregation of the BC1 and BC1_S1 LB plants originating from both J_T3_Tnt1a_2 and J_T3_Tnt1a_5 suggests the presence of a dominant mutation within a gene affecting flowering time, possibly an autonomous pathway gene, caused by the retrotransposon element. It appears that the mutation was homozygous in the T2 generation of plant 5, but only heterozygous in plant 2. See figure 4.71 for an explanation of the genetics occurring in the LB Jessy lines.

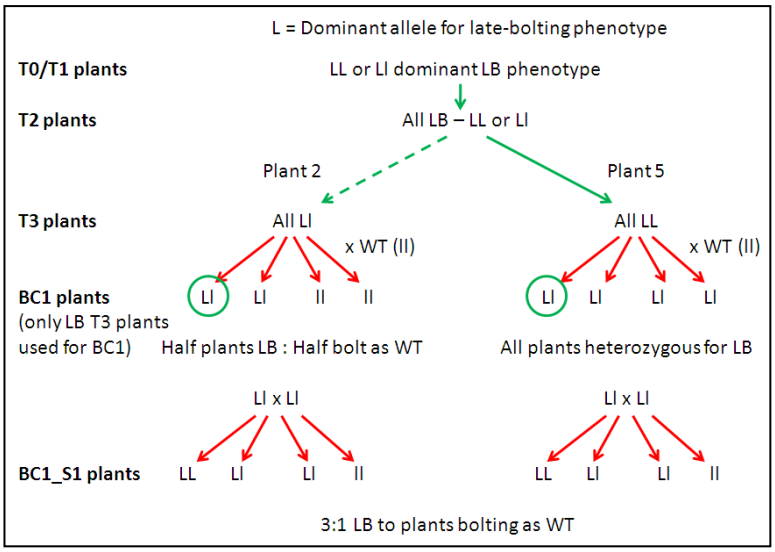


Figure 4.71 – Schematic illustrating the potential segregation of the LB mutation in the Jessy lines

To identify the genes into which the *Tnt1* retrotransposon elements have inserted themselves into in the two LB Jessy lines, inverse PCR (iPCR) and genome walking methods have been attempted, section 5.3.2.2 discusses these experiments in detail.

4.8 Wild lettuce diversity set

4.8.1 Introduction

This experiment involved screening a diversity set of wild lettuce to identify naturally occurring allelic variation which contributes to a LB phenotype. The diversity set was screened with the intention of identifying LB lines which could be compared at a molecular level with lines bolting earlier, any genetic variation identified, particularly in the flowering time genes targeted, could possibly be contributing to the variation in bolting time. If the same gene is polymorphic in two or more LB lines in the diversity set this would be good evidence that it may be linked to the LB phenotype.

The second aspect of this work was to identify LB lines that could be recommended to plant breeders, these lines could be integrated into breeding programs creating new germplasm, potentially with a more robust bolting phenotype. The LB lines identified would be crossed to Saladin as part of this project and act as a pilot study for any future work. Any sequence variation in the candidate genes will be characterised.

The diversity set included wild lettuce from three species; *L. serriola*, *L. saligna* and *L. virosa*, figure 4.72 shows examples of each species.

Figure 4.72 – Examples of wild lettuce species, *L. virosa*, *L. saligna* and *L. serriola.*

L. saligna and *L. serriola* look very similar, but can be distinguished from each other by leaf shape and inflorescence type, see figure 4.72.

4.8.2 Materials and Methods

The wild diversity set was made up of 120 lines; 47 *L. serriola* lines, 42 *L. saligna* lines and 31 *L. virosa* lines, for details of the origin of the lines sown see table 1.3 and Appendix I table A1. Six seeds per line were sown, transplanted and randomised as described in section 2.1.2.1 and 2.1.2.3. Many wild lettuce species are biennial plants and require a period of vernalisation when at the seedling stage to induce flowering, e.g. winter conditions. A vernalisation treatment of 4 weeks at 4 °C has shown to have the most appropriate effect on lettuce seedlings (Prince *et al*., 1979). Seedlings with 3-4 true leaves were treated in this way prior to being transplanted in the glasshouse.

4.8.3 Results and discussion

The results for each species will be described separately;

4.8.3.1 *L. serriola*

Six seeds per *L. serriola* line were sown and scored for number of days to bolt under an artificial LD photoperiod of 16h in the glasshouse. Of the 47 *L. serriola* lines obtained three did not germinate. The remaining 44 lines bolted between 71-140 days, a spread of 69 days. Line 5093 bolted earliest after an average of 71 +/-0 days, the latest line, W58, bolted after an average of 122.6 +/-4.39 days. The plants in the *L. serriola* population bolted at an average of 96.4 days, this is highlighted on figure 4.73 with an orange dotted line. Five lines, which bolted after 113 days or more, were classified as LB compared with the population as a whole, highlighted in green.

Two lines bolting at approximately 96.4 days, the population average, are highlighted in orange, a further two lines, which bolted earliest are highlighted in red. These nine lines were rescored in the glasshouse under a natural LD photoperiod to verify their bolting phenotype, each was confirmed (data not shown). The five LB lines were used to cross to *L. sativa* cv. Saladin, the results of this are described in section 4.8.4. Leaf material was also collected from each of the nine lines for subsequent molecular screens.

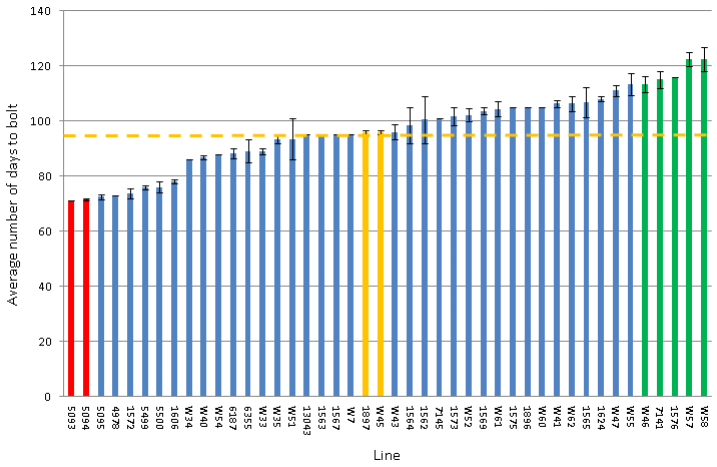


Figure 4.73 – *L. serriola* diversity set, average number of days to bolt. The orange dotted line indicates the average number of days to bolt for the total population. The orange bars represent lines which bolted after a similar number of days to the total population, the red and green bars highlight lines which were classified as EB or LB respectively.

4.8.3.2 *L. saligna*

Six seeds per *L. saligna* line were sown and scored for number of days to bolt under an artificial LD photoperiod in the glasshouse. 42 *L. saligna* lines were obtained for this screen, all germinated and produced bolting data. A range of 57 days in number of days to bolt was observed, the earliest plants bolted after 92 days and the latest after 149 days, see figure 4.74. The plants making up the *L. saligna* population bolted at an average of 114.9 days, this is highlighted with an orange dotted line in figure 4.74. The earliest line to bolt; W90, bolted after an average of 97.4 +/-1.12 days and the latest; line W81, bolted after an average of 145.7 +/-1.67 days. Five

lines which bolted at an average of 132 days or more, which was 17 days later than the average number of days taken to bolt for the population as a whole, are highlighted in green in figure 4.74. Two lines with an average number of days to bolt similar to the average number of days to bolt for the whole population are highlighted in orange with a further two lines which bolted earliest of all, highlighted in red. These nine lines, were rescreened under a natural LD photoperiod, each line displayed the same bolting phenotype, (data not shown). The LB plants were crossed with Saladin and leaf material was collected for subsequent molecular analysis.

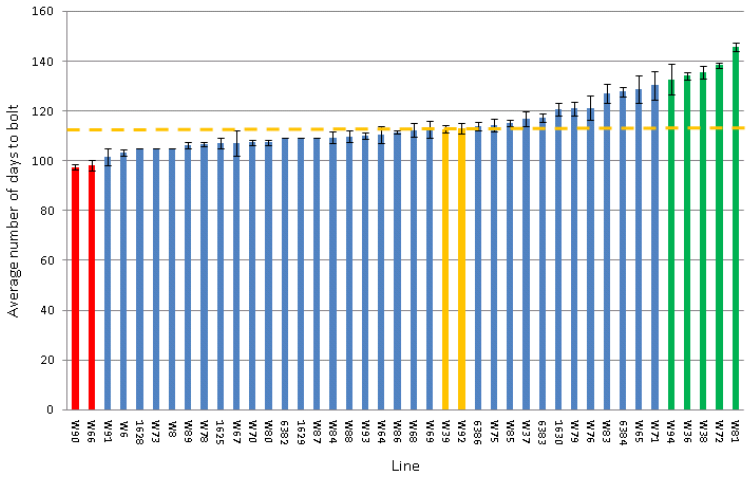


Figure 4.74 – *L. saligna* diversity set, average number of days to bolt per line. The orange dotted line indicates the average number of days to bolt for the total population. The orange bars represent lines which bolted after a similar number of days to the total population, the red and green bars highlight lines which were classified as EB or LB respectively.

4.7.3.3 *L. virosa*

Six seeds per *L. virosa* line were sown and scored for number of days to bolt under an artificial LD photoperiod in the glasshouse. 29 of the 31 *L. virosa* lines obtained for this screen, germinated and were screened. A range of 51 days to bolt was observed, the earliest plant to bolt, bolted after 85 days compared to the latest plant which bolted after 136 days, see figure 4.75. The average number of days to bolt for the entire *L. virosa* population was 93.5 days, and is highlighted in figure 4.76 with an orange dotted line. The earliest line to bolt was line W23 which bolted after an

average of 86.2 +/-0.73 days, the latest line to bolt; line W27 bolted after an average of 99.5 +/-3.38 days. However, of particular interest are the six lines which did not bolt throughout the duration of the experiment, highlighted in green in figure 4.76. These plants showed no signs of bolting after 170 days, and therefore were classified as LB. Two lines bolting after an average number of days similar to that of the average number of days to bolt for the entire *L. virosa* population are highlighted in orange and the two lines bolting the earliest are highlighted in red.

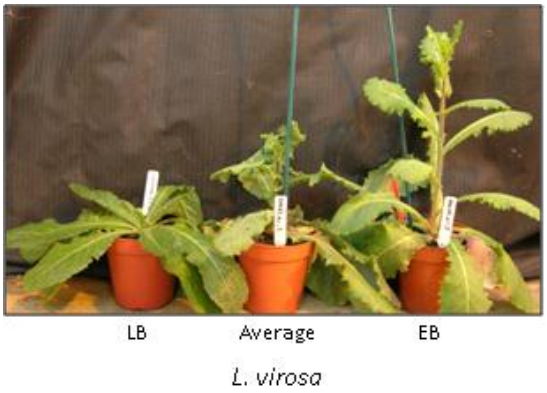


Figure 4.75 – Examples of the range of bolting time observed within *L. virosa* lines.

All 10 of these lines were rescreened as before under a natural LD photoperiod. Plants representing the six LB lines which did not bolt were subjected to a vernalisation treatment of eight weeks at 4ºC to see if a longer period of cold would cause these lines to bolt. All ten of the lines displayed the same bolting phenotype under a natural LD photoperiod, data not shown. Once again the six LB lines did not bolt, this time after being left for 200 days.

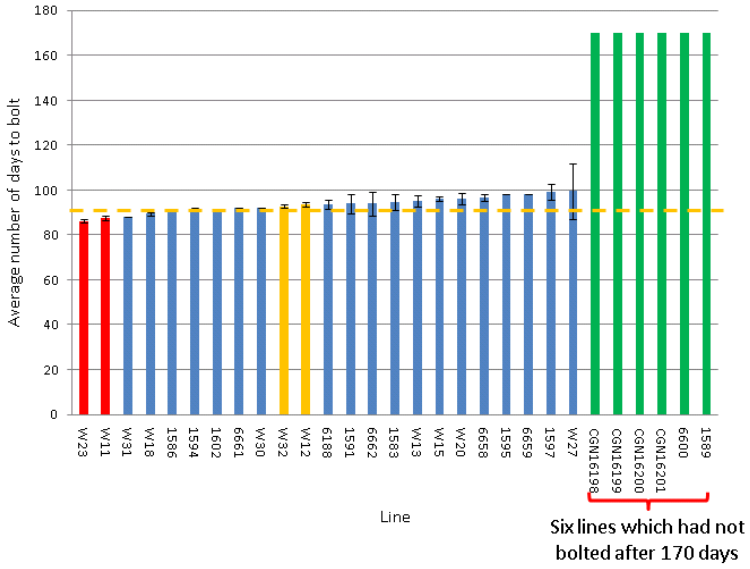


Figure 4.76 - *L. virosa* diversity set, average number of days to bolt. The orange dotted line indicates the average number of days to bolt for the total population. The orange bars represent lines which bolted after a similar number of days to the total population, the red and green bars highlight lines which were classified as EB or LB respectively.

4.8.4 Summary

Table 4.9 summarises the lines from each species that were rescored, of which leaf material was collected for future molecular screens. The LB lines from each species, highlighted in green are those which will be crossed with Saladin.

Species	Bolting phenotype	Line	Species	Bolting phenotype	Line	Species	Bolting phenotype	Line
L. serriola	Early (37 days)	5093 5094	*L. saligna*	Early (62-69)	W90 W66	*L. virosa*	Early (49-57)	W23 W11
	Average (60-72)	1897 W45		Average (73-83)	W39 W92		Average (61)	W32 W12
	Late (80-105)	W46 7141 1576 W57 W58		Late (89-115)	W94 W36 W38 W72 W81		Late (>200)	CGN16198 CGN16199 CGN16200 CGN16201 6600 1589

Table 4.9 – Summary of lines displaying a range of bolting phenotypes from each of the three wild lettuce species screened. The numbers in brackets indicate the bolting time ranges for each of the wild lettuce species.

The difference in the average number of days to bolt observed in each of the lines from the three wild lettuce species suggests that this diversity set contains allelic

variation which significantly affects the bolting time. Crossing between the LB wild lettuce plants and Saladin WT was attempted as described in section 2.2.16. Crosses took place in both directions with both the wild line and the Saladin plants acting as the pollen donor. Crossing was extremely difficult; the flowers of many of the wild lettuce varieties, especially *L. virosa* and *L. saligna* opened very early in the morning and so it was extremely difficult to synchronise flower availability for crossing. Other problems included the size of flowers produced by some of the wild lettuce lines and the relatively small amounts of pollen produced.

Only two of the *L. serriola* lines crossed with Saladin produced any seed. Using Saladin pollen as the donor, lines W47 and W57 successfully set seed. Relatively little seed was produced and unfortunately on sowing this seed did not germinate. Wild lettuce is notoriously difficult to cross with cultivated varieties. De Vries (1990) attempted many crosses between the three wild species and *L. sativa*, he had limited success but did manage to generate some fertile hybrids from crosses with *L. serriola*.

One of our collaborators, Rijk Zwaan®, the Dutch plant breeding company have had success crossing wild lettuce varieties with cultivated lettuce varieties. The LB wild lettuces that have been identified from this work have been recommended to them as potentially interesting lines to follow up, with a view to incorporating them into their breeding programs.

4.9 Overall summary

LB lines have been identified from a number of different lettuce populations. Table 4.10 summarises the findings from this research. Five homozygous LB lines have been identified from a *L. sativa* cv. Larissa EMS population. These lines have been verified under a number of different environmental conditions. The Saladin EMS test and main populations produced 52 lines which were LB after one or two rounds of BC to *L. sativa* cv. Saladin WT. These lines will be rescreened to verify lines them as homozygous for the LB phenotype. A further two LB lines from a *L. sativa* cv. Jessy population have been identified and will be rescreened to confirm the lines are homozygous. A number of LB wild lettuce accessions were identified; *L. serriola* (5), *L. saligna* (5) and *L. virosa* (6).

All of the LB lines will be screened at the molecular level in an attempt to identify the mutation(s) causing the LB phenotype.

Population		No. LB lines identified	Stage of development
Larissa EMS		5	Homozygous LB mutant lines after BC and selfing. Lines to be integrated into breeding programs
Saladin EMS		52	Homozygous mutant lines to be confirmed from 52 lines which have been BC1/BC2 and selfed
Lines with *Tnt1* element		2	Homozygous mutant lines to be confirmed after BC and selfing.
Wild diversity set	*L. serriola*	5	LB lines to be integrated into breeding programs to create new germplasm
	L. saligna	5	
	L. virosa	6	

Table 4.10 – LB lettuce lines identified from EMS mutagenised, *Tnt1*containing and wild lettuce lines

CHAPTER 5

Identification of Allelic Variation in Late Bolting Lettuce Lines

5.1 Introduction

This chapter deals with the identification of genetic variation which exists between WT and the LB lines identified in each population. The aim of this work was to attempt to identify mutations which are causing the LB phenotype. Co-segregation of a particular polymorphism in a candidate gene with a LB phenotype in the BC and selfed lines from each population would be strong evidence that the polymorphism within that gene was causing the bolting phenotype.

One approach taken was to look for polymorphisms occurring in the eight targeted lettuce flowering time genes that are homologues of known flowering genes that had been characterised in *Arabidopsis*. The isolation of these genes from lettuce is described in Chapter 3. A suitable screening method to identify polymorphisms in these genes in our LB lettuce lines had to be chosen. Many sequence based technologies were available to us and were tested to decide how appropriate each method was, taking in to consideration the level of throughput, the robustness and reproducibility of the results obtained and the costs involved. In addition to this candidate gene approach it was decided that mutations that occur in non-targeted flowering time genes and potentially novel flowering time genes involved in flowering in lettuce should be investigated. For this an approach based on transcriptome sequencing was adopted. This chapter is therefore set out in two sections; firstly screening the LB lines for mutations in the target flowering time genes, and secondly screening the LB lines for mutations occurring in all expressed genes, including non-targeted and novel flowering genes by transcriptome

sequencing. The second section also contains the results obtained from screening the Jessy lines containing the tobacco *Tnt1* retrotransposon element.

Material collected from each of the LB lines identified in the original screens of each population was initially used to search for mutations, if a mutation was found in this material, then the co-segregation of the mutation could be followed through each subsequent generation of the line, including material collected from the BC and selfed plants which still had a LB phenotype, and from those which did not have a LB phenotype. The presence of the mutation in a LB line that had been BC and selfed but absence in a sister plant also BC and selfed that no longer had the LB phenotype, would be evidence that this mutation was possibly causing the LB phenotype.

5.2 Materials and Methods

5.2.1 Screening LB lettuce lines for polymorphisms within the lettuce target flowering time gene homologues

5.2.1.1 Light cycler

The Light cycler® 480 (Roche diagnostics), was initially used to look for sequence variation occurring within the lettuce flowering time genes. PCR reactions were set up as follows; 5 µl of 480 SYBR green light cycler master mix (Roche diagnostics, Cat. No., 4707516001) was mixed with a 1 µl pool of forward and reverse PCR primers at a concentration of 10 µM. The amount of $MgCl_2$ added per reaction was optimised for each set of PCR primers used, 1 µl of 20 ng/µl genomic DNA was added as template for each reaction, the reaction volume was made up to 10 µl using sterile distilled water. The PCR primers used were designed with specificity to the lettuce flowering time gene in question, a fragment of approximately 500 bp - 1 kb in

size was shown to be optimal for this method. Cycling conditions for each PCR reaction were as follows; an initial pre-incubation period of 10 m at 95 ºC was followed by 40 cycles of 95 ºC for 30 s, 30 s at a temperature deemed appropriate for the melting temperature of the PCR primers and 30 s at 72 ºC. A melt curve was then constructed using the following conditions; 95 ºC for 1 m, 40 ºC for 1 m, with an increment of 2 ºC per cycle, and 65 ºC for 15 s for 25 cycles. This method was also carried out on pools of DNA, the pools were made up of two to six individual genomic DNA samples, each at a final concentration of 20 ng/µl.

5.2.1.2 TILLING of the Larissa population for mutations within *LsFT*

A 96 well plate of genomic DNA containing samples at 50 ng/µl was set up. The plate was comprised of 60 Larissa M2 and M3 samples from plants which had been screened for days to bolt, see section 4.5.2.1 for details. Of the 60 Larissa samples, four replicates of the LB lines, 2, 164 and 307 were put onto the plate. The plate also contained two replicates of Larissa WT DNA along with three replicates of DNA samples extracted from the control Larissa plants generated from seed treated with water instead of EMS. The plate also contained genomic DNA samples extracted from Saladin WT, Jessy WT, along with two LB Jessy lines The plate also contained samples from a selection of wild lettuce lines, LB lines and lines which bolted after an average number of days from *L. serriola*, *L. saligna* and *L. virosa* species were included. Three water 'blanks' were also included on the plate, for details of the layout of the samples used in this experiment see Appendix IX table A54. An equal volume of 50ng/µl Larissa WT DNA was added to each of the 96 samples. Primers that amplified 1-1.3 kb of sequence were designed to cover the full length of the *LsFT* gene, ensuring an overlap of 150 bp existed between each primer pair. The

TILLING experiments were performed at the Genomics Centre, John Innes Centre (JIC), Norwich, UK; a detailed description of this method can be found in section 2.2.15.

5.2.1.3 Confirmation of polymorphisms identified by TILLING in Larissa lines 199 and 185

100 M3 seeds representing Larissa line 199, 20 M3 seeds representing Larissa line 185 along with 20 WT seed were sown and transplanted in the glasshouse. The plants were grown under a natural LD photoperiod and were scored for days to bolt. Leaf material from each plant was collected and genomic DNA was extracted using the CTAB method described in section 2.2.1. The DNA extracted was screened for the presence of the SNP identified in the TILLING experiment. Line 199 was screened using the restriction enzyme *MboI* (New England BioLabs Ltd., Cat. No. R0147L), a 509 bp PCR product using primers designed specifically to *LsFT* sequence surrounding the mutation was amplified from each sample. 20 µl of PCR product was digested with 1 U of *MboI* in a 50 µl reaction volume at 37 ºC for 1 h, the products were analysed on a 1 % agarose gel. Line 185 was also screened using primers designed specifically to *LsFT*. A PCR product of 998 bp was amplified using FT_TILLING2_F x FT_TILLING2_R, the product was purified, as described in sections 2.2.3 and 2.2.5, the product was then sequenced to verify the presence of the mutation.

5.2.1.4 Analysis of the expression of *LsFT* in Larissa line 199

M3 plants representing the Larissa line 199 were grown in the glasshouse, genomic DNA was extracted from leaf material using the CTAB method, as described in 2.2.1

when the plants had six to eight true leaves. The DNA was genotyped for the mutation identified in the TILLING screen as described in section 5.2.1.3. Two plants which were homozygous for the mutation along with two WT plants were transferred to growth cabinets and were grown under a LD photoperiod for a further week. At this point leaf material was collected every two hours, over a 48 h period. Leaf material was also collected on a weekly basis, from the time at which the plants had 3-4 true leaves (~3 weeks after sowing), until the plants bolted, the plants used in this developmental time course experiment were grown in the glasshouse. *AtFT* is expressed at the end of a LD (Samach *et al*., 2000), and so material was collected 16 h after subjective dawn.

Total RNA was extracted from each of the samples using the Z6 method, described in section 2.2.6, 5 µg of total RNA was DNAse treated, before 1 µg was used as template to synthesis cDNA. PCR primers for Real-Time analysis were designed to the *LsFT* gene sequence, and Real-Time PCR was carried out as described in section 2.2.11.

5.2.1.5 PCR screen of target genes in interesting LB lettuce lines

Genomic DNA from each of the LB Larissa and Saladin lines was extracted using the CTAB method described in section 2.2.1. PCR primers were designed to amplify the full genomic sequence of each of the genes targeted in this project; PCR reactions were optimised to produce single PCR products which were purified and were sequenced directly using the PCR primers, along with a number of internal primers designed specifically to the PCR product. Where possible PCR reactions were carried out using KOD Taq polymerase, as described in section 2.2.3. However to maximise the size of the products amplified, especially for products of 3 kb in size or

more and where products were difficult to amplify, KOD Xtreme™ Hot Start DNA polymerase (Novagen, Cat. No., 71975-3) was used. Each Xtreme™ reaction took place in a 50 µl volume, 1.5 µl of 15 µM forward and reverse PCR primers was added to 25 µl of 2x reaction buffer, 10 µl of 10 mM dNTPs and 1 U of KOD Extreme™ Hot Start DNA polymerase. 1 µl of genomic DNA was used as the template for each reaction, the DNA used was either at 100 or 1000 ng/µl. The reaction volume was made up to 50 µl with sterile distilled water. The cycling conditions used were as follows; an initial denaturation step of 94 ºC for 2 m was followed by 35 cycles of 98 ºC for 10 s, 30 s at an annealing temperature deemed to be appropriate for the melting temperature of the primers designed and an extension at 68 ºC for 1 min/ kb of product expected, before a final extension cycle of 4min at 68 ºC. The primers used for each of the PCR reactions and those used in the sequencing reactions of each of the products generated can be found in Appendix XX figure A1.

5.2.1.6 Analysis of the expression of the target genes in LB Larissa lines compared with Larissa WT

To investigate whether the expression of the flowering time gene homologues identified in the LB mutant Larissa lines was affected throughout the plant's development in comparison with Larissa WT, semi quantitative PCR was carried out. Leaf material was collected from each of the Larissa WT and the LB Larissa lines 2, 164 and 307 at the developmental time points, including samples taken pre- and post-bolt specified in table 5.1.

Line	Day	Developmental stage of plant	Days to bolt
WT	40/47 55/63/69/78	Pre-bolt Post-bolt	52
2	40/47/55/63 69/78	Pre-bolt Post-bolt	65
164	40/47/55 63/69/78	Pre-bolt Post-bolt	56
307	40/47/55/63/69 78	Pre-bolt Post-bolt	75

Table 5.1 – Tissue collected from LB Larissa lines for developmental time course experiment

Total RNA was extracted from the leaf material, 5µg was treated with DNAse and cDNA was synthesised from 1µg of each treated sample. Standard PCR reactions using KOD Taq polymerase as described in section 2.2.3 was carried out using 0.5µl of cDNA as template. Each PCR reaction was optimised to ascertain a suitable number of cycles where the amplified product was in the logarithmic phase. This is important to enable the presence of any differences in gene expression between the lines to be easily identified. The PCR primers used for each reaction can be found in Appendix II.

5.2.1.7 Screening the Larissa LB lines for mutations affecting photoperception

LB Larissa lines, along with Larissa WT, were screened for potential mutations in photoreceptor genes or in the downstream photoperception pathways by measuring the length of seedling hypocotyl after 10 days growth. 12 seeds representing Larissa WT and seeds from each of the three LB Larissa lines (2, 164 and 307) which had been BC and selfed were sown in F2 compost and were stratified at 4 ºC for 72 h in the dark. They were then transferred to an environmental control chamber (Percival Scientific Inc., Cat. No., 30LED), and were subjected to continual far-red light, blue light, white light (at 1 $\mu Mm^{-2}s^{-1}$) or dark conditions for 10 days, after which the hypocotyl lengths were measured.

5.2.2 Screening LB lettuce lines for polymorphisms in non targeted and novel flowering time genes.

5.2.2.1 Screening LB Larissa lines 164 and 307 for SNPs using Illumina Genome Analyser II technology

5.2.2.1.1 Preparation of material

To identify SNPs within non-targeted and/or novel flowering time genes in lettuce, two LB Larissa lines (164 and 307), along with Larissa WT were sequenced at the transcriptome level. This method used Illumina Genome Analyser II technology (Illumina Inc.).

Material was collected from different lettuce tissue at various developmental stages of growth as detailed in table 5.2, to try to ensure that genes expressed in certain tissues and/or a certain time points were all included in the transcriptome sequencing reactions. The LB lines screened were S2_BC1_164_E1 and S2_BC2_307_B4. A depth of coverage of five to ten fold for the Larissa transcriptome was desired to provide confidence in the data generated. Total RNA was extracted from the samples using Z6 buffer as described in section 2.2.6.

Tissue collected	Details	Resuspension volume (µl)
Leaf material (Developmental time course)	Pool of material from 5 plants grown in LDs in the glasshouse, collected once a week at ZT16, starting when plants have 3-4 true leaves and ending prior to plants bolting	30
Leaf material (Diurnal time course)	Pool of leaf material from 5 individual plants/line grown in LDs in growth cabinets, collected at ZT 0.5, 4, 8, 12, 15, 16 and 19.5	30
Apex material (Pre/Post bolt)	Pre bolt apices collected 2 days prior to bolting from 12 individual plants/line. Post bolt collected 7 days after bolting from 3 individual plants/line all grown in LDs in glasshouse	10
Flowers	2 fully opened flowers collected from 2 individual plants/line grown in LDs in glasshouse	40
Vernalised material	All aerial parts of 3 seedlings/line subjected to 4°C for 4 weeks.	40
Imbibed seed	4 seed/line subjected to 24h of continuous light at 20°C	20
Etiolated seedlings	All aerial parts of 4 seedlings/line grown in the dark at room temperature, collection took place 7 days post cotyledon emergence	20
Juvenile seedlings	All aerial parts of 4 seedlings/line grown under SDs, collected 1 day and 4 days post cotyledon emergence	20

Table 5.2 – Tissue collected from LB Larissa lines for transcriptome sequencing

The total RNA samples were quantified on the nanodrop and 5 µg was DNAse treated as described in section 2.2.6. Using an RNA 6000 Pico Kit (Agilent Technologies, Cat. No., G2941-90161), the quality of the RNA samples was tested on a bioanalyser 2100 (Agilent Technologies), following the manufacturer's guidelines. RNA from each collection was pooled together as follows to create a 10 µg stock in a final volume of 50 µl; 2.5 µg of RNA extracted from the diurnal time course, 2 µg from a combined pool of RNA extracted from pre-bolted developmental and vernalised leaf material, 2 µg from a combined pool of RNA extracted from

juvenile and etiolated seedlings, 2 µg from a combined pool of RNA extracted from pre- and post-bolt apices, 0.5 µg of RNA extracted from flower material, 0.5 µg of RNA extracted from imbibed seed material and 0.5 µg of RNA extracted from post-bolt leaf material.

5.2.2.1.2 Preparation of cDNA library for analysis

The cDNA libraries were prepared by the Genomics Centre at Warwick HRI using the mRNA-Seq 8-Sample Prep Kit (Illumina® Inc., Cat. No. RS-100-0801), following the manufacturer's guidelines. The first step was to purify mRNA from the Total RNA extracted, this was done by purifying the poly-A mRNA molecules using poly-T oligo-attached magnetic beads. The mRNA was then fragmented into small pieces using divalent cations under elevated temperature and first round cDNA was synthesised using reverse transcriptase and random primers, before double stranded cDNA was generated by removing the strand of mRNA and synthesising a replacement strand. The cDNA ends were repaired by converting the overhangs into blunt ends using T4 DNA polymerase and Klenow DNA polymerase. An 'A' base was added to the 3' end of the blunt phosphorylated DNA fragments using the polymerase activity of Klenow fragment, preparing the DNA fragments for ligation of adaptors which had a single 'T' overhang at their 3' end. Adaptors were ligated to the ends of the DNA fragments, this was followed by purification on a gel to select templates with a size range of approximately 200 bp; the purified cDNA was enriched using PCR primers that annealed to the end of the adaptors. The sample quality, size and concentration were verified using a bioanalyser.

5.2.2.1.3 Sequencing of the cDNA libraries

Paired-end sequence reads of 76 bases per read, were generated by sequencing using a Genome Analyser II (Illumina® Inc). WT and mutant samples were sequenced in a single lane each in the same Illumina flowcell.

5.2.2.1.4 Analysis of data (performed by Dr. Jay Moore, Warwick HRI)

Two *Lactuca* EST libraries were accessed from the CGP2 website, the EST library constructed from *L. sativa* material; LACT_SATI.CSA1, and the EST library comprising data collected from five *Lactuca* species; LACT_5CDS.CSA1. The libraries were used as reference sequences for the alignment of the transcriptome 76 bp paired end reads obtained by sequencing mRNA extracted from Larissa WT samples. The transcriptome sequences were aligned to the ESTs making up the database using the BOWTIE SAMtools (Li *et al*., 2009) software. WT consensus sequences were then constructed based on the alignments made, the contigs generated, covering regions of the Larissa WT transcriptome were then were filtered to remove ambiguous base calls. The transcriptome sequence reads obtained for the LB samples (lines 164 and 307), were then aligned to the WT consensus sequences, again using SAMtools before being filtered to remove ambiguous calls in aligned mutant sequences. Sequence differences were identified between the WT and LB lines by lining up the sequences, the depth of coverage for each individual nucleotide was also calculated for both the WT and LB sequences, this data provided an indication of the confidence level for each potential sequence difference. The blast2go program (www.blast2go.org/start_blast2go) was then used to annotate each of the ESTs containing a SNP with a functional description relating to the region of the genome that the EST aligns to, these annotations are known as a set of GO terms.

The program performs BLAST searches against public and private databases. The GO terms generated relate to the function of each protein that is most highly conserved to each EST; e.g. flowering, meaning that searching for sequence differences in genes involved in processes such as flowering were more easy to identify. The sequence differences observed were ranked based on the depth of coverage and were then sorted based on the GO terms assigned using a set of terms relevant to this project (e.g. flower, flowering, flowered, temperature, vernalisation, cold, photoperiod, photoperiodic, photoperiodism, autonomous, temperature, photoreceptor, phytochrome and cryptochrome).

5.2.2.1.5 Analysis of interesting polymorphisms

Polymorphisms identified in genes involved in flowering time or with any of the relevant GO terms were verified using PCR. Primers were designed based on the EST sequence containing the polymorphism of interest, see Appendix II for details. PCR fragments of approximately 500 bp were amplified using the standard KOD Taq polymerase method as described in section 2.2.3, the products were purified and sequenced see section 2.2.5 and 2.2.10, to confirm the presence or absence of the polymorphism identified from the transcriptome sequencing.

5.2.2.2 Analysis of LB Jessy lines containing the *Tnt1* retrotransposon element

5.2.2.2.1 PCR screen of target genes

A PCR screen of genomic DNA extracted from the LB Jessy lines J_T3_Tnt1a_2 and J_T3_Tnt1a_5 using the CTAB method described in section 2.2.1, was carried out using the conditions as described in section 2.2.3. Genomic DNA extracted from Jessy WT was also screened as a control. Each of the lettuce flowering target genes

identified was screened for the possible presence of the retrotransposon element. Products obtained for each reaction were analysed on 1 % agarose gels. The absence of any product or a product, of an unexpected size, could indicate the presence of the *Tnt1* element within the gene.

5.2.2.2.2 Inverse PCR and Genome Walking

Two methods were attempted to identify genes in to which the *Tnt1* retrotransposon elements had inserted themselves in the LB Jessy lines J_T3_Tnt1a_2 and J_T3_Tnt1a_5. Firstly inverse PCR (iPCR) was attempted as described by Lucas *et al*., (1995). Genome walking was also carried out using primers designed to the 3' and 5' ends of the *Tnt1* retrotransposon sequence. The genome walk method is described in section 2.2.7.

5.3 Results

5.3.1 Screening LB lettuce lines for polymorphisms within the lettuce target flowering time gene homologues

5.3.1.1 Analysis of PCR fragments from the target genes in LB Larissa lines using light cycler technology

The light cycler® 480 was initially tested as a potential method to screen the LB lettuce lines for sequence variation in the candidate genes that may be present in the Larissa EMS population and the wild lettuce diversity screen. Melt curves for each of the PCR products amplified are generated, any differences observed between the melt curve obtained from WT and mutant lines would suggest the presence of a sequence difference within the fragment amplified.

Initially the method was tested using genomic DNA extracted from leaf material collected from Larissa WT, the LB Larissa mutant lines, 2, 164 and 307, and some of the non-LB Larissa lines 61, 99, 195 and 303, along with *L. serriola* line 1576, *L. saligna* line W72 and *L. virosa* line 6600 which had LB phenotypes compared to other lines from the same species. A 331 bp PCR product was amplified covering *LsFT* exon 1 to exon 2 from each line using primers FT_EST_1_F and FT_EST_1_R, see section 2.2.3 for method used. As the melt curve profile produced for each sample was identical it suggested that there were no sequence differences within this fragment in the LB Larissa lines 2, 164 and 307 compared with WT and the other Larissa lines which did not display a delayed bolting phenotype. However the melt curves produced during the amplification of the PCR product from the three wild lettuce lines are significantly different from those generated from the Larissa lines, see figure 5.1.

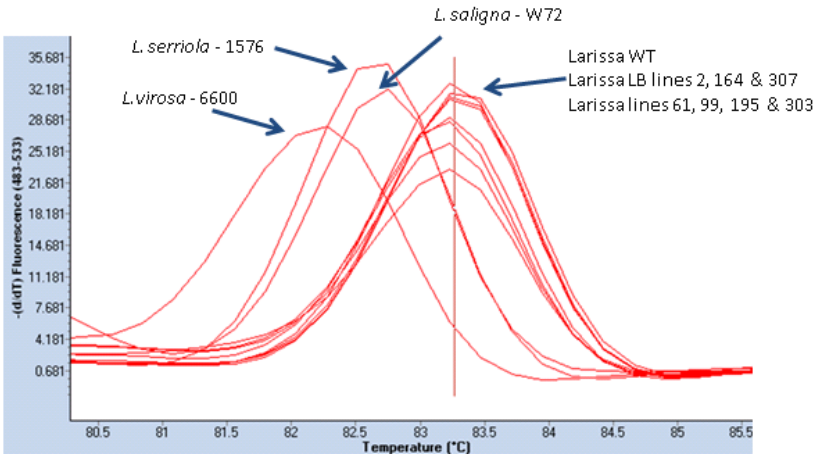


Figure 5.1 – Melt curve produced for an *LsFT* PCR product amplified from Larissa lines and wild lettuce species using light cycler technology. The melt curves produced from the amplification of PCR products from the Larissa lines are all identical, suggesting no sequence variation. Shifts in the melt curves generated from PCR products amplified from the three wild lettuce species suggests the presence of one or more sequence differences.

To investigate whether any sequence differences existed within the 331 bp fragments amplified from the lines tested, the PCR reaction was repeated and the products amplified were purified and sequenced.

Sequence obtained from Larissa WT and the three LB mutant lines confirmed there was no sequence variation present in this region of the *LsFT* gene. No sequence was obtained from the *L. saligna* line W72. However, *L. serriola* line 1576, which produced a melting curve containing a peak at 82.75 ºC; 0.5 ºC earlier than WT contained four sequence differences. The fragment amplified from *L. virosa* line 6600, contained nine sequence differences when compared with Larissa WT see figure 5.1. Interestingly the shift in melt peak obtained from *L. virosa*, (1ºC earlier than Larissa WT), is bigger than that seen with *L. serriola* line 1576 see figure 5.1, suggesting the level of sequence variation is proportional to the shift in melt curve. The position at which the sequence differences were observed along with any resulting amino acid change are summarised in table 5.3.

Nucleotide Position	Gene Region	Line	Amino acid change
91	exon 1	vir 6600	N/C
113	exon 1	vir 6600	T→S
120	exon 1	vir 6600/ser 1576	S→N
141	exon 1	vir 6600	N/C
156	exon 1	ser 1576	N/C
13	intron 1	vir 6600	-
50	intron 1	ser 1576	-
92	intron 1	vir 6600	-
106	intron 1	vir 6600	-
108	intron 1	vir 6600/ser 1576	-
9	exon 2	vir 6600	N/C

Table 5.3 – Sequence differences identified between the Larissa lines and *L. serriola* line 1576 and *L.virosa* line 6600 in a 331 bp PCR product amplified from *LsFT* using the light cycler. N/C = no change in amino acid, - = no amino acid change occurs as the sequence change is within an intron.

Of the four sequence differences seen between Larissa WT and *L. serriola* line 1576, two occur within the first exon of *LsFT*, however neither result in an amino acid change, see figure 5.2. However of the nine changes observed between Larissa and *L. virosa*, four occur within either exon 1 or exon 2, and two of these changes, both in exon 1, result in an amino acid change. The first change at the 91st nucleotide of exon 1, an adenine to thymine substitution results in the amino acid translated

changing from a threonine to a serine. The second nucleotide substitution at exon 1 position 113; a guanine to adenine substitution, results in the amino acid translated changing from a serine to an asparagine, see figure 5.2.

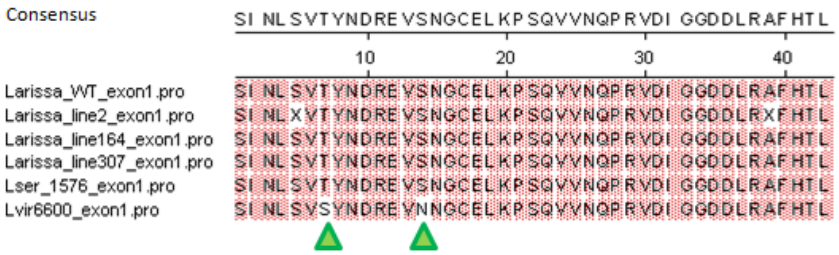


Figure 5.2 – Sequence comparison of Larissa WT and LB lines with *L.serriola* (line 1576) and *L. virosa* (line 6600). Two nucleotide substitutions within the first exon of *LsFT* result in two amino acid changes.

The amount of variation in just 331 bp of *LsFT* between Larissa and the wild lettuce lines is interesting. There are four polymorphisms between Larissa and *L. serriola*, and nine polymorphisms between Larissa and *L. virosa*, two of which cause changes in the protein. To see if the two nucleotide substitutions observed in *LsFT* exon 1 in *L. virosa* line 6600 could be causing the LB phenotype in the line, PCR fragments covering *LsFT* exon 1 were amplified from other lines making up the *L. virosa* diversity set which were not LB and bolted at the same time as the average number of days recorded for the species, lines W32 and W12 (61 days to bolt), and lines bolting early when compared to the rest of the species; lines W23 and W11 (49-57 days to bolt), along with another LB line; line 1589 (>200 days to bolt). All of the products amplified and sequenced contained the same nucleotide changes as observed in *L. virosa* line 6600, as recorded in table 5.3. Therefore the amino acid differences that exist are species specific and are not causing the LB phenotype of lines 6600 and 1589.

To further test the capabilities of the light cycler, pools of DNA were constructed to see if mutations could still be identified when screening a larger number of pooled samples. PCR reactions, using *LsFT* exon 1 to exon 2 primers were tested on pools

of DNA containing Larissa WT DNA spiked with DNA from the LB *L. virosa* line 6600. Pools containing a 1:1, 1:2, 1:3 and 1:4 dilution of *L. virosa* line 6600 to Larissa WT were created. The presence of the mutations within *L. virosa* line 6600 could not be detected in any of the pools.

The original method, using non-pooled samples was repeated using primers designed specifically to amplify single PCR products from *LsFCA*, *LsFLD* and *LsFVE*. The products amplified ranged in size from 700-900 bp. Although single PCR products were amplified the light cycler did not produce consistent data from any of the genes. Although this method of mutation detection produced some interesting results with the *LsFT* gene, there were reasons why it was decided not to use it to screen our LB lines. Firstly each reaction requires a lot of optimisation, this method is not particularly cheap and so reagent costs to optimise reactions along with the inability of the method to detect mutations within pooled samples meant the cost and amount of time required would be substantial. Other reasons why this method was not adopted was that only relatively small PCR fragments could be processed, products up to 1 kb were suggested to be workable and some of the genes targeted in this project are 7 kb or more in size, meaning that a large number of reactions would need to be carried out per gene. The final reason for not pursuing this method was that the reproducibility of results obtained was poor; reactions with a fragment amplified from *LsFVE* produced different melt curves each time it was tested.

5.3.1.2 Analysis of *LsFT* PCR fragments from lettuce lines using TILLING

TILLING was the next method tested to identify polymorphisms within the target genes in the LB lettuce lines. The method used is described in sections 2.2.15 and 5.2.1.3, a 96 well plate containing genomic DNA extracted from the Larissa EMS

population, some of the wild lettuce diversity set lines and the some of the lettuce lines containing the retrotransposon element *Tnt1* was set up and sent to be processed by the Genomics Centre, JIC.

The maximum size of PCR product which could be analysed by TILLING was approximately 1.5 kb, because the *LsFT* gene, including approximately 1.2 kb of 5'UTR and 200 bp of 3'UTR is 3.3 kb, three pairs of primers were designed and optimised to amplify single PCR products with 150 bp of overlap between each set, see figure 5.3 for details.

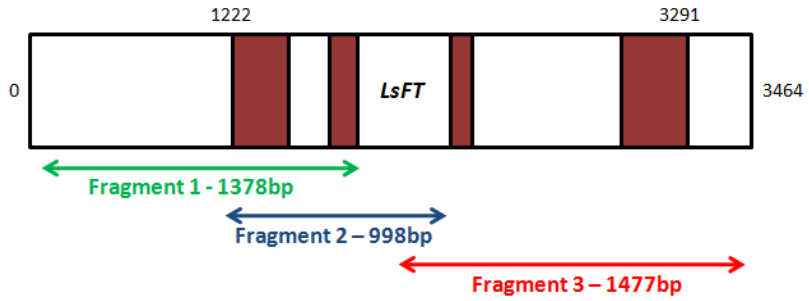


Figure 5.3 – TILLING primers using to screen *LsFT*

TILLING works on the basis that when PCR products amplified from a pool of WT DNA mixed with DNA from a mutant line, are denatured and re-annealed to each other, products containing a sequence mismatch will be produced, these mismatches are then cleaved using the enzyme CEL1. The products obtained are then run on a gel or an ABI sequencer; the lines containing a mutation produce two products compared to one undigested product from WT, see figure 1.9 in section 1.6.3.1. This method worked well for the primers covering the first two fragments of *LsFT*, but despite designing two different sets of primers to cover the final fragment, no data could be generated covering fragment three. This problem was overcome by directly sequencing PCR products amplified from the Larissa population obtained using the primers FT_TILLING3_F and R and comparing the sequence with that of Larissa WT.

Of the 60 Larissa EMS lines screened, the TILLING method identified two which contained potential mutations, these polymorphisms were verified by repeating the relevant PCR, sequencing the products and comparing with the sequence amplified from Larissa WT. This confirmed that lines 199 and 185 both contained a polymorphism within the *LsFT* gene. The mutation in line 199 was present in fragment 1, within the 5'UTR, at position 1101 of the *LsFT* gene sequence, see Appendix VI for full sequence details. A cytosine nucleotide found at position 1101 in Larissa WT is replaced by a cytosine/thymine heterozygous allele at the same position in line 199, see figure 5.4.

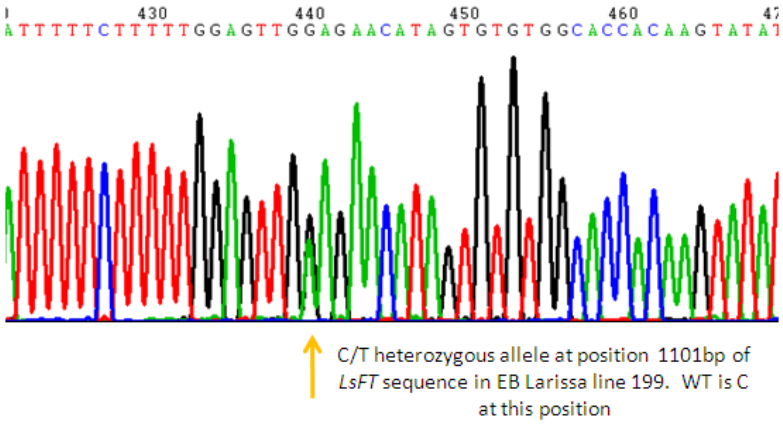


Figure 5.4 – Chromatogram showing a heterozygous allele at position 1101 in *LsFT* in Larissa line 199. WT has a C allele at this position.

The bolting time of line 199 was recorded in the original Larissa EMS population screen carried out under an artificial LD photoperiod in the glasshouse. The plants screened were a combination of M2 and M3 plants, on average line 199 bolted after 66.3 +/-1.27 days compared to WT which bolted after 77.6 +/-1.14 days. Of all the Larissa lines screened, line 199 was the sixth earliest to bolt. The difference in the number of days to bolt was shown to be significant between Larissa WT and line 199 ($p=<0.001$, d.f.=10, l.s.d.=3.04). Figure 5.5 illustrates the bolting phenotype of each of the 199 plants screened during the trial. All eleven bolted earlier than the average

number of days to bolt of the Larissa WT plants (identified with an orange dotted line). Line 199 was represented on the TILLING plate by a sample from plant M2_199_4, highlighted with a green arrow. It is possible that the variation seen in days to bolt for the plants making up line 199 could be caused by the allelic variation observed in *LsFT* at position 1101. M2_199_4 is heterozygous at this position and bolts at the same time as plants M2_199_5, 9 and 10. M2_199_2, 3, 7 and 8 all bolt later than plants 4, 5, 9 and 10, and at a more similar time to the EB WT plants, suggesting that these four plants could be homozygous cytosine, like the WT plants. Three further plants; M3_199_1 and 2 and M2_199_6, bolt earlier than plant 4, these plants could potentially be caused by a homozygous mutant allele (thymine).

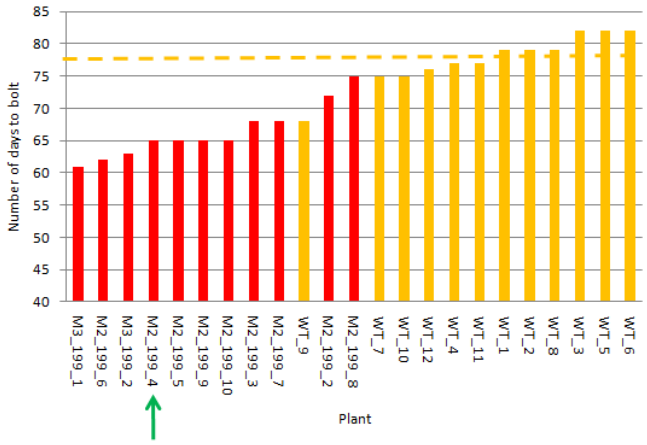


Figure 5.5 – Number of days to bolt for plants from Larissa line 199 compared with WT under an artificial LD photoperiod. The green arrow highlights the plant which was used in the TILLING experiments. The orange dotted line displays the average number of days to bolt for Larissa WT.

As all the M2 seed available was used in the original screen, 100 M3 seeds, (from the same M3 batch of seed used in the original screen, e.g. plants M3_199_1 and 2), were screened, along with 20 Larissa WT plants, for days to bolt. 77 of the seeds sown went on to bolt, however 25 of these plants displayed an undesirable phenotype; some were small and some had very thin elongated leaves (these abnormalities were not observed in the original screen). These phenotypes were

probably caused by the EMS treatment and might have affected the bolting phenotype recorded; therefore these plants were removed from the analysis.

The remaining 52 plants were genotyped for the allele at position 1101 in *LsFT* by digesting a 509 bp PCR product containing the mutation. The PCR product was generated using FT_MboII_199_F and FT_MboII_199_R primers, the product was then digested with the restriction enzyme *MboII*. The presence of a mutant allele at position 1101 in *LsFT* means an *MboII* site is generated. Plants which were homozygous for the WT cytosine allele at position 1101 did not cut and therefore a single band of 509 bp was present. Plants containing a mutant thymine allele produced two products, of 327 and 182 bp and heterozygous plants produced three bands, one at 509 bp representing the cytosine allele and two at 327 and 182 bp representing the presence of the mutant thymine allele. Of the 52 plants; three contained a cytosine homozygous allele at position 1101 as seen in WT plants, 26 were heterozygous and 23 were homozygous for the mutant thymine allele. WT plants bolted after an average of 49 +/-0.61 days. Mutant 199 plants with a homozygous cytosine allele bolted after an average of 44.7 +/-1.2 days (4.3 days earlier than WT), plants with a heterozygous genotype and plants homozygous for the mutant allele, bolted after a similar length of time with an average of 43.4 +/-0.33 and 43.7 +/-0.43 days respectively (see figure 5.6).

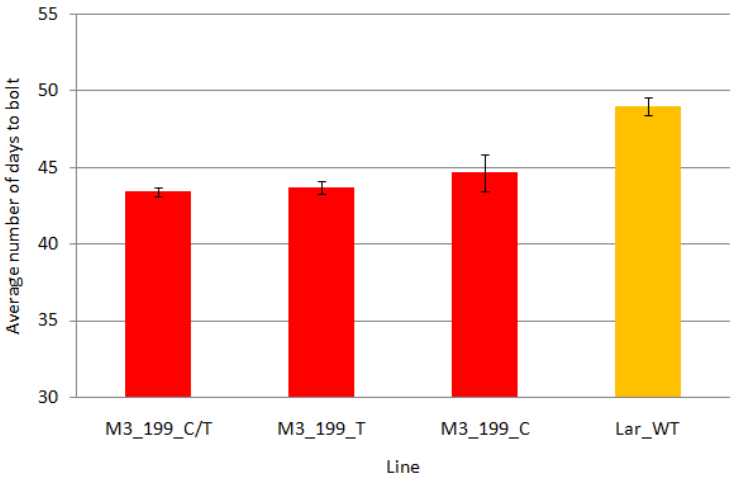


Figure 5.6 – Rescreen of Larissa line 199 compared to WT. Each of the plants was genotyped for the allele at position 1101 in *LsFT*. The average number of days to bolt per genotype did not significantly vary, each genotype resulted in an EB phenotype.

It would appear from this work that the mutation in line 199, at position 1101 in the 5'UTR of *LsFT* is not causing the EB phenotype. Therefore another mutation(s) within line 199 must be causing the EB phenotype observed.

To confirm that the mutant allele was not affecting the expression of the *LsFT* gene in line 199 real-time PCR was carried out. Plant material was collected every 2h over a 48h period from plants homozygous for the mutant allele and from Larissa WT plants. cDNA synthesised from this material was used as the template for the real-time PCR reactions. Figure 5.7 shows the results of this experiment, the expression of *LsFT* relative to *LsELFα* over a 48h period has been averaged and plotted for both Larissa WT and line 199. A peak in expression of *LsFT* in both WT and mutant samples is seen at ZT12-14, the relative level of expression of *LsFT* in both WT and mutant samples is also very similar. This result suggests that the diurnal expression pattern of *LsFT* is not affected by the SNP at position 1101.

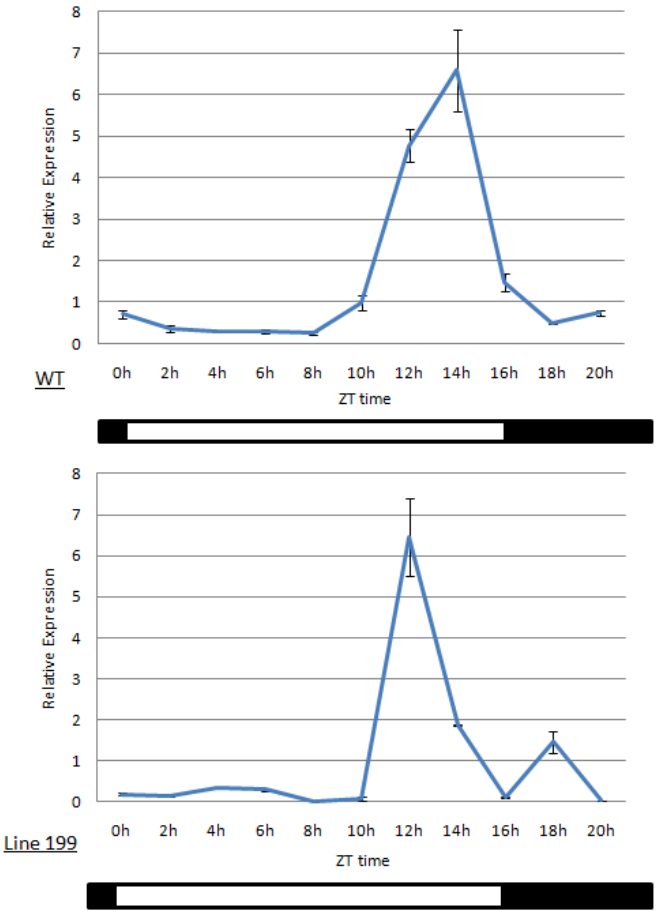


Figure 5.7 – Relative expression of *LsFT* in EB Larissa line 199 compared to WT over a 24h period time course. A peak in expression occurs in both lines at ~12-14h after subjective dawn. Material screened was collected every two hours over a 24h period, plants were subjected to a 16h photoperiod, shown with a white box, the black box indicates when the plants were in darkness.

Developmental time course leaf material was also collected from both Larissa WT and line 199 to see if the mutant allele was affecting the expression of *LsFT* through the plant's development. If the *LsFT* gene is expressed earlier in the development of plants representing line 199 than in Larissa WT plants this could explain the EB phenotype. *LsFT* expression was analysed relative to that of *LsELFα*, samples taken at ZT16 from six time points were analysed; day 33, 40, 47, 55, 63 and 69 days after sowing. Plants representing line 199 bolted after an average of 43.7 +/-0.43 days and showed a significant increase in *LsFT* expression in the sample taken on day 55, approximately 11 days after bolting. Larissa WT plants bolted after an average of 49 +/-0.61 days, a significant increase in *LsFT* expression was observed in the sample

taken on day 63, 14 days after bolting. It appears from this experiment that a similar developmental pattern of *LsFT* expression exhibited exists for both lines. This data further indicates that the mutation at position 1101 in the 5'UTR of *LsFT* is not causing the EB phenotype seen in line 199, see figure 5.8.

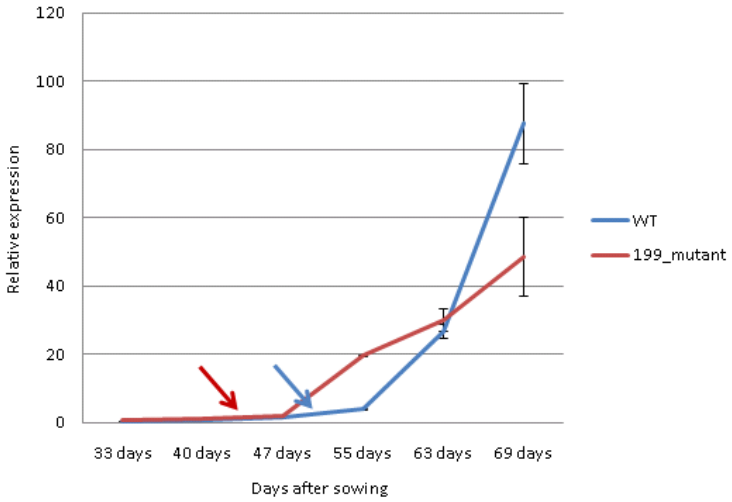


Figure 5.8 - Relative expression of *LsFT* in EB Larissa line 199 compared to WT over a developmental time course. Both plants show an increase in the expression of *LsFT* after the plant has bolted. The arrows indicate the point at which each plant bolted.

Line 199 remains of interest because of its EB phenotype under both artificial and natural LD photoperiods, unfortunately it was not included in the original BC experiments. Now that the bolting phenotype has been shown to be robust, plants are being BC and rescreened. This data is not included in this thesis.

The second mutation identified from the TILLING screen was located in fragment two of Larissa line 185, at position 1393 in the *LsFT* full length gene sequence, see Appendix VI. Position 1393 corresponds to nucleotide 172 of *LsFT* exon 1. The change relates to a guanine nucleotide at position 1393 in WT, being replaced by a guanine/adenine heterozygous allele in line 185. This difference results in the glycine amino acid at position 58 in *LsFT* changing to serine. The presence of the polymorphism in line 185 was verified by amplifying the second *LsFT* TILLING fragment using PCR primers FT_TILLING2_F and FT_TILLING2_R, then

sequencing the product and comparing it to the same region amplified from WT, see figure 5.9.

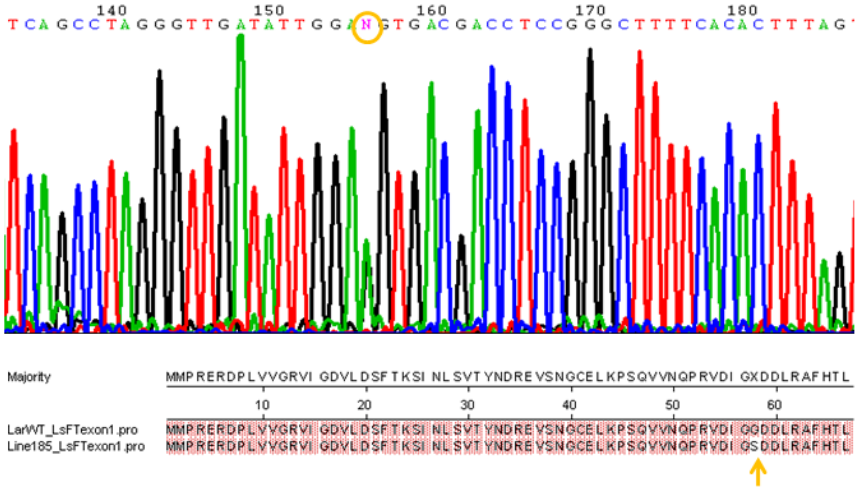


Figure 5.9 – Chromatogram of LB Larissa line 185 and the resultant amino acid change in *LsFT* exon 1 observed. The guanine allele seen at position 172 (highlighted with an orange circle) of *LsFT* exon 1 is replaced by a guanine/adenine heterozygous allele, which results in an amino acid change; glycine to serine (indicated with an orange arrow).

Interestingly line 185 was one of the seven original lines scored as LB under an artificial LD photoperiod, however on rescreening under a natural LD photoperiod the LB phenotype appeared to be lost, see figure 4.16. Figure 5.10A illustrates the original data obtained when line 185 was screened under an artificial LD photoperiod. Nine plants screened were from the M2 generation; all of which bolted significantly later than WT ($p=<0.001$, d.f.=9, l.s.d.=4.9), highlighted with green bars, two M3 plants were also scored, both bolted earlier than WT (blue bars). Four plants from the line were rescreened under a natural LD photoperiod, using seed from the same M3 seed stock used in the original screen, but the plants appeared to have lost the LB phenotype ($p=0.063$, d.f.=3, l.s.d.=5.22), see section 4.5.2.2 for details, and so line 185 was not chosen for BC to WT. The mutation identified in *LsFT* in the TILLING screen was from genomic DNA extracted from plant M2_185_9 (green arrow), which bolted 13 days after WT in the original bolting screen. This plant did not set any seed, so M3 seed was collected from another LB

plant; M2_185_3, which bolted after 102 days, 25 days later than WT. 20 plants from the M3 stock collected from plant M2_185_3 were screened for days to bolt under natural LD conditions. The M3 plants grown from M2_185_3 all bolted significantly later than the WT plants screened (p=>0.001, d.f.=37, l.s.d.=2.03), the M3_185_3 plants bolted after an average of 65.2 +/-0.83 days, on average, 10.7 days later than WT (see figure 5.10B), suggesting the mutation causing the LB phenotype in these plants is homozygous and possibly dominant (because the TILLING screen identified a heterozygous polymorphism in the M2_185_9 plant which is LB).

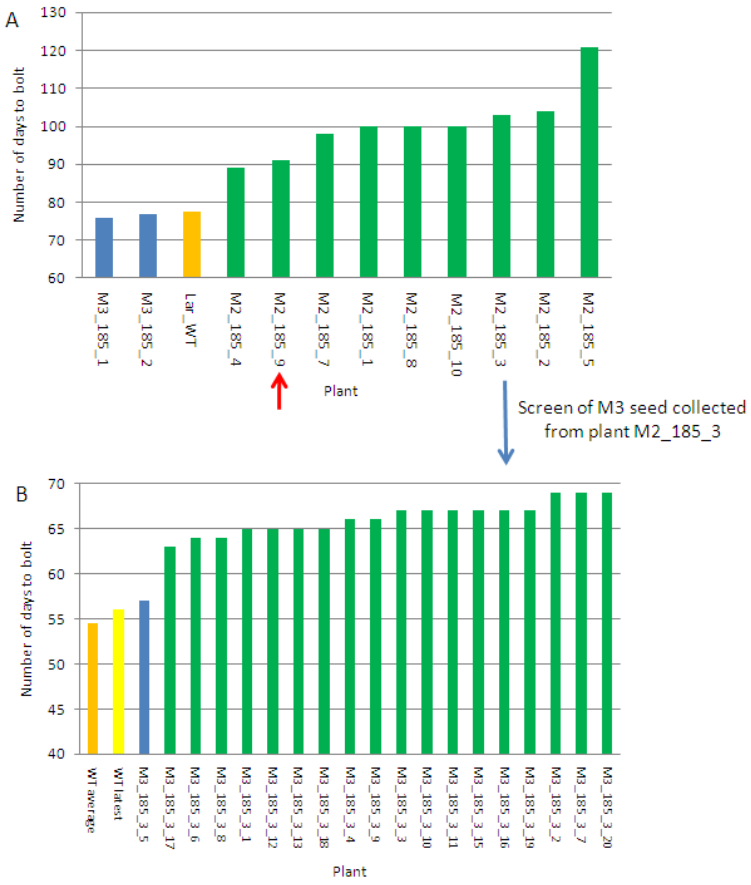


Figure 5.10 – Bolting time of LB Larissa line 185.
(A) - Number of days to bolt recorded for LB line 185 in original screen under an artificial LD photoperiod, all M2 plants screened are LB. The M3 plants bolt as WT. The red arrow identifies the plant used in the TILLING screen.
(B) - Number of days to bolt; M3 population rescreen under a natural LD photoperiod. The seed stock screened was that collected from the LB M2_185_3 plant.
Green bars indicate LB plants and blue bars indicate plants bolting as WT.

Leaf material was screened for the mutation at *LsFT* position 1393 from each of the individual plants scored, in figure 5.10B. None of the samples analysed contained the mutation identified in plant M2_185_9. It appears that the mutation which was present in this M2 plant was not present or has been lost in M3 seeds collected from other M2 plants. Therefore the LB phenotype of Larissa line 185 is not being caused by the mutation observed in exon 1 of *LsFT* in plant M2_185_9, however this is still an interesting line as the LB phenotype of the line does appear to be robust. Further analysis of this line has begun, but will not be discussed in this thesis.

The remaining samples screened using the TILLING method, included those extracted from the mutant lines containing a *Tnt1* retrotransposon element; the Jessy lines screened and the Saladin WT samples contained no polymorphisms, meaning the *LsFT* gene is conserved across the butterhead lettuce cultivars and Saladin. However sequence variation in the wild lettuce lines, including the previously identified sequence variation from the work described in section 5.3.1.1 using the light cycler, was again observed using this method. Although the TILLING experiment picked out mutations within the wild lettuce lines, mixing the samples with Larissa WT DNA meant that many fragments were generated due to the large amount of variation between the wild species and Larissa. This makes it difficult to identify how many products are produced with the CEL I digest. A better approach would have been to make a pool of each LB wild lettuce line together with a line that bolts with an average number of days for that species; a method known as Ecotilling (Comai *et al.*, 2004)

TILLING proved to be a very efficient method of screening the relatively small Larissa EMS population. TILLING is designed, as explained in figure 1.9, to be used on pools of DNA, but this was not attempted due to the small population size

screened. However should the Saladin population be screened using TILLING in future work, then pooling of DNA samples would be a sensible idea. The other target genes were not screened using TILLING because of the significant costs encountered for the labeled primers and the general running costs.

5.3.1.3 PCR screen of target genes in interesting LB Larissa and Saladin lettuce lines

It was decided that a more robust method of screening the LB lettuce lines was required. As full length sequence of the candidate genes in lettuce had been obtained, and screening the LB lines using light cycler and TILLING technologies has proved unsuccessful or inappropriate, it was decided that the LB lines would be screened by amplifying each of the candidate genes by PCR and sequencing the products. The products amplified from each of the LB Larissa and Saladin lines were compared directly with sequence amplified from Larissa WT and Saladin WT plants.

Initially Larissa WT was used as the template DNA in optimising each of the PCR reactions. Primers were designed with the intention of amplifying products of as large a size as was possible, all reactions where products of over 3 kb were expected were optimised using KOD Xtreme™ Hot Start DNA polymerase as opposed to KOD. KOD Xtreme™ has been shown to successfully amplify larger PCR products, especially those over 3 kb. Primers were also designed to include around 1 kb of 5'UTR of each of the genes targeted. Each PCR reaction was further optimised, by altering the concentration of the template DNA, modifying the annealing temperature or by increasing the number of PCR cycles, to produce single products of a suitable intensity that purification of the product would result in the recovery of a concentration of DNA required for sequencing reactions. Table 5.4 shows a

breakdown of the PCR fragments generated and screened. Figure A1, Appendix XX, illustrates each target gene and highlights the whereabouts of each of the sequencing primers used to screen the PCR products. *LsFT* was not screened using this method in the LB Larissa lines as it had been analysed in section 5.3.1.2, using TILLING methods and direct sequencing. However the LB Saladin lines were screened for polymorphisms within *LsFT* using the PCR screening method, the three primer pairs used for the TILLING work were used to amplify single products which were subsequently sequenced and analysed, see Appendix II for details of sequencing primers used.

Gene	PCR primers	Template (ng)	PCR details	Coverage (bp)
FLK	FLK_fulllength_extreme_F x FLK_fulllength_extreme_R	100	KOD Xtreme™, a.55ºC, e.5m, 35c	6-5038
FLD	FLD_fulllength_extreme_F x FLD_fulllength_extreme_R	100	KOD Xtreme™, a.57ºC, e.4m, 35c	66-3956
FVE	FVE_extreme_F x FVE_comp_R	100	KOD Xtreme™, a.55ºC, e.5m, 35c	70-5249
FCA	FCA_extreme x FCA_intron3_R	100	KOD Xtreme™, a.57ºC, e.4.5m, 35c	109-4672
	FCA_intron3_F x FCA_exon13_R	1000	KOD, a.54ºC, e.1m, 30c	4237-6515
	FCA_exon9_F x FCA_extreme_R	1000	KOD, a.55ºC, e.40s, 30c	5725-7905
FPA	FPA_fulllength_F x FPA_intron1_R	1000	KOD, a.52ºC, e.2m, 30c	521-3899
	FPA_exon1_F x FPA_EST1_R	1000	KOD, a.55ºC, e.40s, 30c	1665-4861
	FPA_intron1_F1 x FPA_exon5_R2_new	100	KOD, a.55ºC, e.20s, 30c	4563-7848
LD	LD_5UTR_F x LD_RegionI-II_R2	1000	KOD, a.55ºC, e.40s, 30c	378-2714
	LD_RegionI-II_F x LD_fulllength_R	100	KOD, a.52ºC, e.1m, 30c	1909-4530
	LD_seqint11 x LD_Dec09_R3	100	KOD, a.55ºC, e.1m15s, 30c	4096-7268
	LD_Reg3_F x LD_Reg3_R	100	KOD, a.60ºC, e.40s, 30c	6935-8307
FY	FY_extreme_F x FY_extreme_R2	100	KOD Xtreme™, a.56ºC, e. 5m, 35c	30-4628
	FY_intron3_F x FY_exon14_R	100	KOD Xtreme™, a.58ºC, e.4m, 35c	2337-6251
	FY_intron11_F2 x FY_exon16_R	100	KOD, a.56ºC, e.40s, 30c	5725-7637
	FY_exon13_F x FY_extreme_R	100	KOD, a.50ºC, e.1m, 30c	6047-9305

Table 5.4 – PCR primers and reaction conditions used to screen the LB lines for mutations. a. = annealing temperature, e. = extension time, c = cycles

Having successfully optimised the PCR reactions and sequenced each of the candidate gene fragments amplified from Larissa WT, the same products were amplified from the LB Larissa lines, 2, 164 and 307. The template genomic DNA used in the PCR reactions was extracted from leaf material collected from individual LB plants that were screened in the first glasshouse trial, M2_2_10, M3_164_1 and M2_307_3, these plants bolted after an average of 93 +/-3.04, 91.33 +/-2.71 and 90.7 +/-2.62 days respectively, compared with WT which bolted after an average of 77.6 +/-1.15 days. Ten LB Saladin lines were also screened using this method; five of the lines screened were from the 'test' population, and genomic DNA was extracted from leaf material collected from the following LB plants grown in the first haygrove trial; M2_38_1, M2_64_2, M2_75_4, M2_124_4 and M2_131_1 which bolted after 152, 152, 166, 181 and 166 days respectively, compared with WT which bolted after 125 days. These five lines from the 'test' population were chosen because each produced plants with a healthy phenotype and each line was successfully BC twice to WT and selfed. The other five Saladin lines screened originated from the 'main' population; lines 335, 372, 808, 843 and 844, each of these lines bolted as WT in the M1 generation; but in the M3 population each plant representing these lines was LB, see sections 4.6.1.6 and 4.6.1.9, suggesting the mutation affecting these lines was recessive. These lines are potentially of most interest in terms of the lettuce flowering genes that have been targeted in this project; the mutations in *FT* and the autonomous pathway genes causing late flowering in *Arabidopsis* are recessive. It is therefore possible that the LB phenotype in these five lines is caused by a similar recessive mutation in one of these genes. As all M3 plants screened for these lines were LB, M4 seed, which should be homozygous for the mutation causing the LB phenotype, was collected and sown, DNA was extracted from leaf material collected

from the following lines; M4_335_4_D, M4_372_1_B, M4_808_4_D, M4_843_1_C and M4_844_1_B to use in the PCR screen.

The strategy undertaken was to identify sequence differences in the original M2/M3/M4 lines screened and to then follow the mutation through the subsequent generations of the LB lines, e.g. in material collected from plants at each of round of BC and selfing, to discover whether the mutation is segregating within the line and determine whether it is associated with the LB phenotype. In some cases several putative lettuce homologues may exhibit high levels of similarity to a known flowering time gene, the identification of a polymorphism that affects flowering time in one of them would identify that gene as a true orthologue.

Eight genes (*FT* and the seven autonomous pathway genes) were screened in the ten LB Saladin lines and the sequences obtained were directly compared to the sequence obtained from Saladin WT. No sequence variation was identified between the LB lines and Saladin WT within any of the lettuce genes targeted. This means that an EMS generated mutation occurring within a non-targeted flowering time gene, is causing the LB phenotype in the ten LB Saladin lines.

Sequence data obtained from the LB Larissa lines 2 and 164 also contained no differences when compared to Larissa WT. However two sequence differences were identified in line 307 when compared to Larissa WT, they were identified in the genes *LsFVE* and *LsLD*.

The mutation within *LsLD*, identified in the PCR product amplified using the primers LD_5UTR_F x LD_RegionI-II_R2, occurred at position 2022, see Appendix VI for full gene details, within the second intron. The polymorphism results in a guanine to adenine substitution, the change was confirmed using three sequencing primers covering both strands of DNA, LD_RegionI-II_F, LD_intron1_F and LD_RegionI-

II_R2. The chromatogram in figure 5.11 shows the sequence comparison between WT and line 307 generated using the sequencing primer LD_RegI-II_R2. The mutation identified does not have an effect on the coding sequence or amino acid composition of the gene, nor does it create a potential alternative splice site.

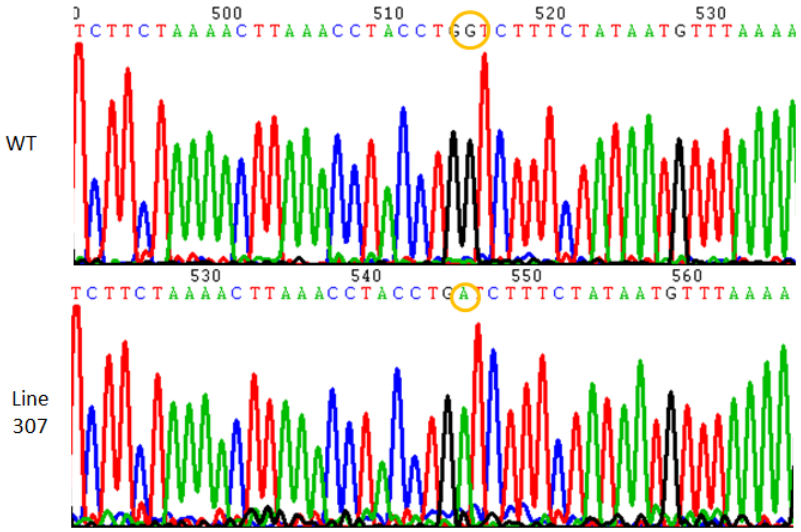


Figure 5.11 – Chromatograms highlighting the sequence difference observed between WT and LB Larissa line 307 in *LsLD*.

The second mutation identified in line 307 is within *LsFVE*, at position 3244 in the full length gene sequence, see Appendix VI for full gene sequence details. The sequence change was identified within a PCR product amplified using FVE_extreme_F x FVE_comp_R and corresponds to a guanine nucleotide in Larissa WT being substituted with an adenine nucleotide in line 307. Position 3244 is located 11 bases into intron 5 of *LsFVE*. As seen with the change observed in *LsLD*, the polymorphism in *LsFVE* does not alter the coding sequence or amino acid composition of the gene, nor does it create an alternative splice site. The mutation was sequence verified using four different primers; FVE_exon2-4_F, FVE_exon4-6_F, FVE_exon7-8_R and FVE_intron3_F, covering the region. See chromatogram in figure 5.12 for a sequence comparison between WT and line 307 generated using the sequencing primer FVE_exon7-8_R.

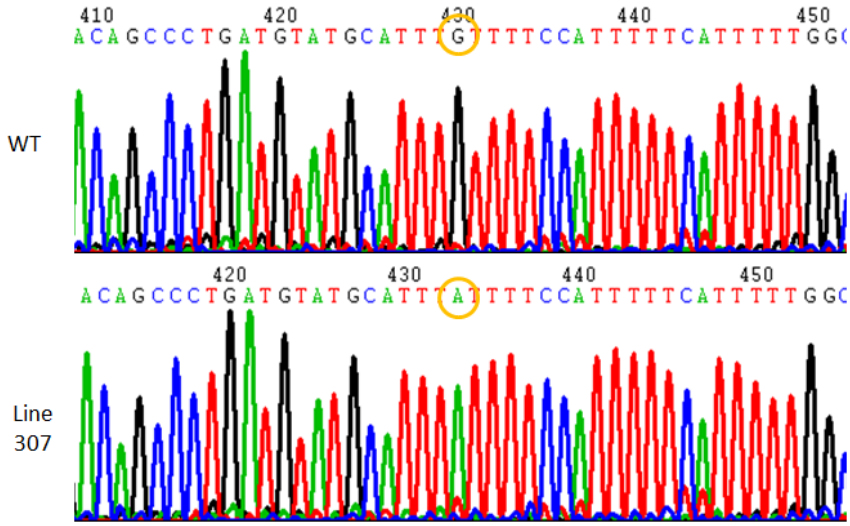


Figure 5.12 – Chromatograms highlighting the sequence difference observed between WT and LB Larissa line 307 in *LsFVE*.

To investigate whether the polymorphisms identified in line 307 are present in the homozygous mutant 307 lines which have been subjected to two rounds of BC and a self fertilisation and still maintain their LB phenotype, genomic DNA was extracted from 307_BC2_S1_B4_4, a plant which bolted after 88 days compared to WT which bolted after an average of 62.4 +/-1.48 days. Genomic DNA was also extracted from another plant representing line 307, 307_BC2_S1_B2_2, which had been through two rounds of BC and a self fertilisation, but which no longer displayed a LB phenotype, it bolted after 57 days. The rationale behind this was that if either polymorphism was present in the LB plant, but not present in the line bolting as WT after BC and selfing, then the effect of the mutation could be investigated in more detail with the hypothesis that it was causing the bolting phenotype observed.

The PCR products containing the mutations within *LsLD* and *LsFVE* were amplified from the two BC and selfed plants using the conditions described in table 5.4. Neither plant contained either of the mutations that had been identified in the original LB M2 plant, this means that the polymorphisms identified in both *LsLD* and *LsFVE* were not contributing to the LB phenotype. However, the LB phenotype must be

caused by a mutation that was present in the original M2 line 307 plant and is still present in plant 307_BC2_S1_B4_4, but has been lost through BC in plant 307_BC2_S1_B2_2; this mutation is not within any of the genes targeted in this work though.

Sequencing purified PCR products generated from the target flowering genes in the LB Larissa and Saladin lines proved to be a good method for obtaining conclusive and reproducible results. Full length gene sequence were obtained from WT and each of the LB lines, very few regions proved difficult to sequence, occasionally repeat sequence within introns did provide difficulties for sequencing methods. It is interesting to see that there are very few mutations within our target genes in these lines, this is not entirely unexpected as the mutation rate caused by the EMS treatment in other crop species has been shown to be as low as 1 per 200 kb in pea (Dalmais *et al*., 2008), and 1 per 300 kb in tomato (Minoia *et al*., 2010). A rough mutation rate observed in the LB Larissa line 164 after 2 rounds of BC was calculated using the transcriptome sequencing data discussed in section 5.3.2.1 of approximately 1 per 12.5 kb. It would be expected that the mutation rate would be lower in the Saladin 'test' population due a lower concentration of EMS used, and even lower in the 'main' Saladin population as the seeds were subjected to only 18 h in 0.1 % EMS. Although the data produced was definitive and of a high quality the method is labour intensive, time consuming and reasonably expensive; a more efficient, high throughput method would be beneficial for future screening of the lettuce populations. It was interesting to note that very little sequence variation within the target genes was noted between Diana, (the cultivar in which the BAC library was constructed), Larissa and Saladin.

Unfortunately, targeting flowering time genes and attempting to identify mutations within them in LB lines did not produced many results to follow up, and the polymorphisms that were identified in *LsFT* in lines 199 and 185 through TILLING, and in *LsFVE* and *LsLD* in line 307, have been shown to not have a causative effect on the LB phenotype.

5.3.1.4 Expression of the target genes in the LB lines

From the sequencing work described in section 5.3.1.3 it has been shown that there are no mutations present within the coding sequence of each of the lettuce flowering time gene homologues. To investigate whether a mutation within a gene (or microRNA) which acts upstream of one of the target flowering time genes may affect its expression thus causing delayed bolting, semi quantitative PCR was performed to analyse the expression of the genes over a developmental time course. Leaf material was collected from Larissa WT and LB lines 2, 164 and 307, throughout the pre- and post-bolt stages of development, see table 5.1 for details of the samples taken. RNA was extracted and cDNA was synthesised from each of the samples before PCR conditions were optimised for primers designed to each of the genes targeted. This work tested the hypothesis that the effect of a mutation within a gene directly regulating one of the lettuce flowering genes targeted would have a significant effect on the expression of this gene and thus may affect bolting time. If expression levels appear to be affected using semi-quantitative PCR then real time PCR could be carried out for more detailed analysis.

The semi-quantitative PCR results obtained from pre- and post- bolt samples collected from Larissa WT and the LB Larissa lines for all eight of the target genes

showed no obvious differences in gene expression were observed when comparing WT with LB Larissa samples.

The expression of the four *LsFLC* gene homologues identified in lettuce (see section 3.3.13) was also analysed. *FLC* is a strong repressor of flowering in *Arabidopsis*. The rationale behind testing the *LsFLC* genes was that should one of the four genes be an *AtFLC* orthologue then any change in expression levels may not only be down to a mutation occurring within the *LsFLC* gene but may be potentially caused by a mutation within a gene acting upstream of *LsFLC*, which include the seven autonomous pathway genes targeted. This is important to bear in mind as the lettuce homologues of the *Arabidopsis* target genes identified in this project may not be the functional *Arabidopsis* gene orthologues. It may also be that the genes making up the autonomous pathway have not all been characterised in *Arabidopsis.*

Semi quantitative PCR did indicate that there were some differences in the level of expression of each of the *LsFLC* genes 1, 2 and 4 in the pre and post bolt material analysed, no differences were identified in the expression of *LsFLC* gene 3. Real Time PCR analysis confirmed these differences. There is an increase in the expression of *LsFLC* gene 1 in the post bolt samples of the three LB Larissa lines screened, see figure 5.13, the peak in expression occurs at ZT8 in the post bolt leaf material collected on day 78 (highlighted with orange arrows). The data obtained from WT is less clear, although it is unlikely that the differences in expression observed in the LB lines is having an effect on the bolting phenotype. It is unlikely that this lettuce gene is the functional orthologue of *AtFLC* as an increase in the expression of the *LsFLC* gene would be expected in the pre-bolt samples if it was causing the LB phenotype observed in these lines.

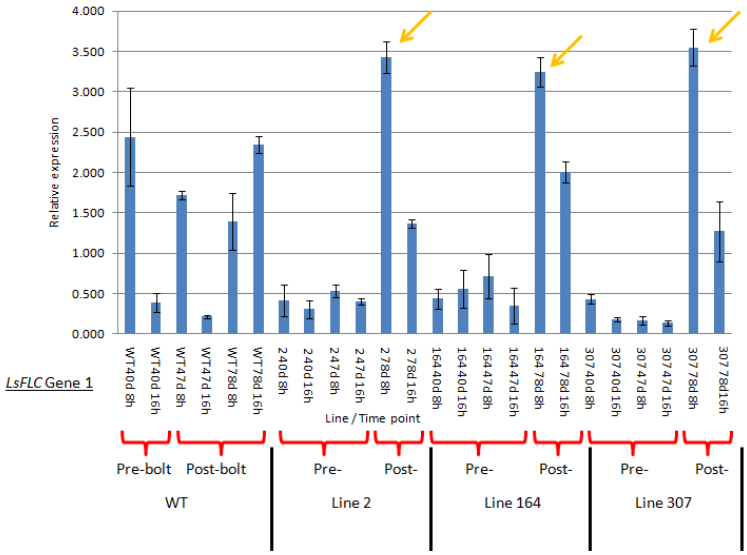


Figure 5.13 – Relative expression of *LsFLC* gene 1 in pre- and post-bolt material collected from WT and LB Larissa lines

The level of expression seen for *LsFLC* gene 2 appears to be diurnally controlled, in each of the samples analysed, irrespective of whether the material was collected pre- or post-bolt, *LsFLC* gene 2 is always most highly expressed in the sample collected at ZT8 (blue bars) in comparison to the sample taken at ZT16 (purple bars) on the same day, see figure 5.14. There is no great difference in the level of expression of *LsFLC* gene 2 in pre- or post-bolt material collected from Larissa WT and the LB lines.

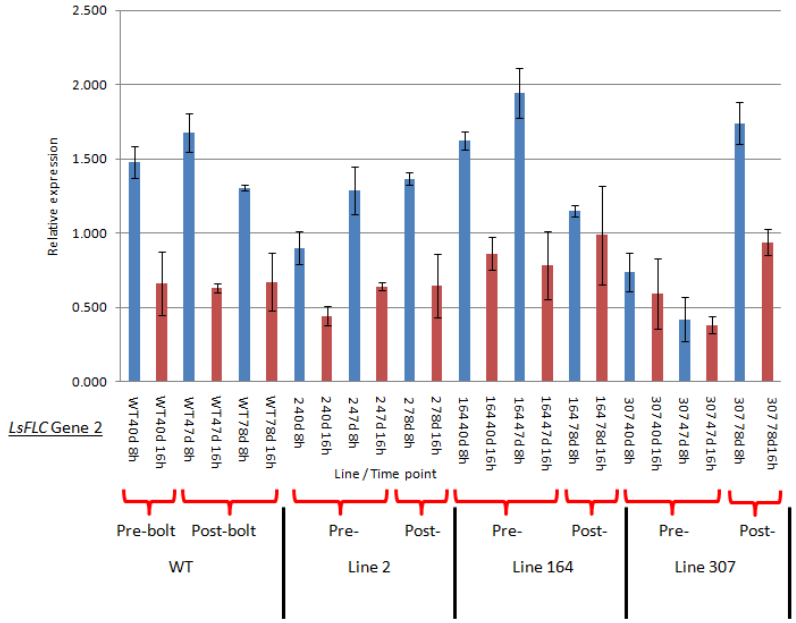


Figure 5.14 – Relative expression of *LsFLC* gene 2 in pre- and post-bolt material collected from WT and LB Larissa lines. The blue bars highlight material collected 8h after dawn, the red bars indicate material collected at 16h after dawn.

Real-time PCR was also carried out to look at the expression of *LsFLC* gene 3, this confirmed what was seen with the semi-quantitative PCR, that the level of expression was constant in all samples at all time points (data not shown).

LsFLC gene 4 is expressed most highly in the pre-bolt samples collected from each of the LB lines and Larissa WT. The level of relative expression of *LsFLC* gene 4 in comparison with *LsELFα* is much higher than the expression of *LsFLC* gene 1 or 2 at any developmental stage, see figure 5.15. The level of expression seen in *LsFLC* gene 4 does seem to drop as the plant develops, however it decreases more slowly in the LB lines than in WT where it is negligible by day 47. This slower decline in *LsFLC* gene 4 expression levels may be the cause of the delayed bolting in these lines, the presence of the *LsFLC* repressor delaying the floral transition, alternatively it may be that *LsFLC* gene 4 is only expressed in the vegetative stage of development and not following the floral transition. In any case it would appear that the

expression of *LsFLC* gene 4 is linked to the vegetative stage of lettuce plant development.

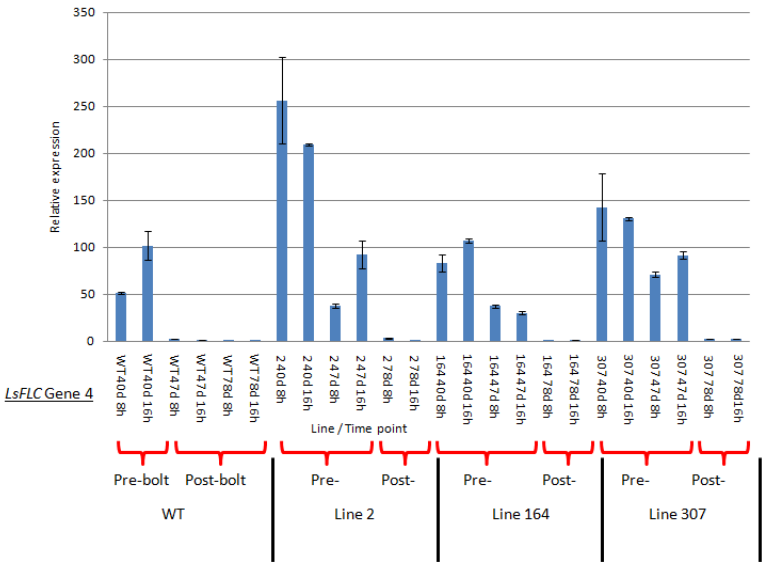


Figure 5.15 – Relative expression of *LsFLC* gene 4 in pre- and post-bolt material collected from WT and LB Larissa lines

There is no difference in the level of expression in *LsFLC* gene 4 in Larissa WT compared to the LB lines, this suggests that as with *LsFLC* genes 1, 2 and 3, there is not a mutation within the *LsFLC* homologues identified affecting the bolting phenotype of the LB lines. However as the timing of *FLC* gene 4 expression is altered in the LB lines it is possible that a mutation within a gene acting upstream is affecting the level of expression of this *LsFLC* gene homologue and this may be causing the LB phenotype of these lines.

5.3.1.5 Testing LB Larissa lines for potential mutations in photoreceptors

Photoreceptors such as *PHYA*, *PHYB* and *CRY2* regulate flowering time and photoperiodic responses in *Arabidopsis.* In particular, null mutations in *PHYB* and *CRY2* cause early flowering and delayed flowering respectively. Such mutations are readily detected because they also result in altered hypocotyl elongation in specific

wavelengths of light, such as red light for *phyB* mutations or blue light for *cry2* mutations (Ahmad *et al.*, 2002).

To identify whether any of the Larissa LB lines were affected in their PHYB or CRY2 signal transduction pathways they were screened under continuous red or blue light, as well as under dark and continuous white light conditions. The lines screened are detailed in table 5.5; seedlings representing LB Larissa lines, 2, 164 and 307, which had been BC twice and selfed were used in this experiment, along with seedlings representing lines 2, 164 and 307, which had been through the BC and selfing process but no longer displayed a LB phenotype. The length of 12 seedlings was measured after being subjected to 10 days of each light treatment. The hypocotyl lengths of the seedlings for each line under each set of conditions were averaged, see table 5.5.

Line	Bolting phenotype	Hypocotyl length (cms)			
		Dark	Blue	Red	White
WT	N/A	10.25	7.64	6.3	0.31
2_C1	LB	11.33	8.84	6.76	0.4
2_D3	LB	9.23	7.29	5.94	0.43
2_G2	LB	10.04	7.69	5.96	0.53
2_D1	As WT	8.86	5.68	5.26	0.61
164_E1	LB	9.63	7.22	6.61	0.67
164_E3	LB	9.98	7.39	6.95	0.78
164_A3	As WT	11.03	8.17	6.92	0.61
307_B4	LB	9.46	7.7	6.86	0.83
307_E2	As WT	10.6	7.96	7.19	0.98
307_E4	As WT	10.18	8.59	7.59	0.89

Table 5.5 – Details of Larissa lines screened for mutations within lettuce photoreceptor genes and the average hypocotyl length scored under each set of light conditions.

Figure 5.16 shows data collected from one representative LB individual, from each of the LB Larissa lines along with WT under each of the different light conditions. Although some small differences in hypocotyl length were observed in this screen, a much more significant difference would have been expected if a mutation was

present within the *PHYB* and *CRY2* photoreceptor genes or the downstream signaling pathways.

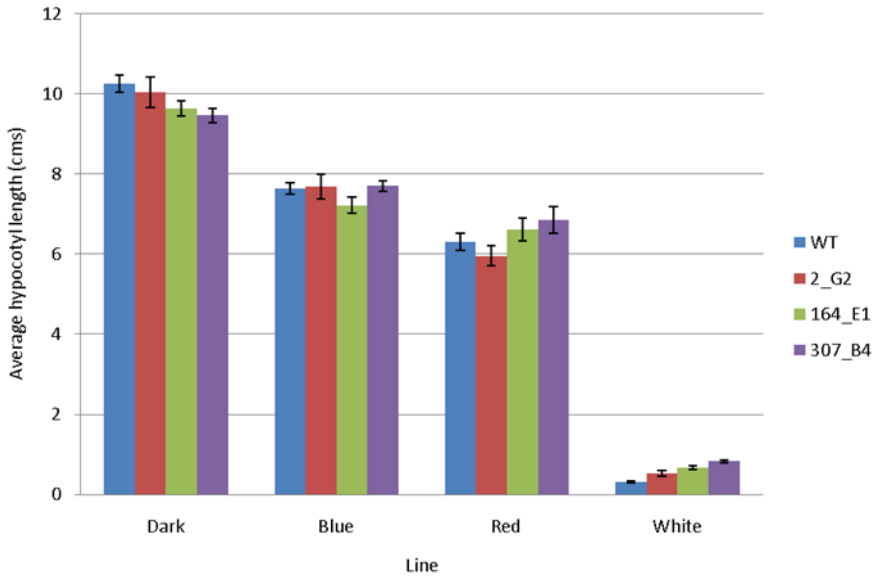


Figure 5.16 – Average hypocotyl length of seedlings from Larissa WT and seeds originating from the LB Larissa lines 2, 164 and 307 which were subjected to dark, blue light, red light or white light conditions for 10 days. No significant difference in average hypocotyl length compared with WT was observed.

This data suggests that there were no mutations present within the *PHYB* and *CRY2* photoreceptor genes or the downstream signaling pathways in the LB Larissa lines 2, 164 and 307.

5.3.2 Screening LB lettuce lines for polymorphisms within non-targeted flowering time genes and novel lettuce

5.3.2.1 Identification of mutations in LB Larissa lines 164 and 307 using Illumina Genome Analyser II technology

The rationale of this experiment was to look for sequence variation (i.e. SNPs induced by the EMS treatment) occurring between the transcriptome sequences of Larissa WT and the LB Larissa lines 164 and 307. The Genome Analyser sequencing run for Larissa WT, LB Larissa lines S2_BC2_164_E1 and S2_BC2_307_B4 generated 33,727,826, 32,253,722 and 46,990,686 76 bp paired-

end reads respectively. The first step in identifying sequence variation between the two LB lines and Larissa WT was to create a reference sequence for Larissa WT. This was done by combining the sequences representing the *L. sativa* cv. Salinas EST database; LACT_SATI.CSA1 and sequences representing five *Lactuca* species (*perennis, sativa, serriola, saligna* and *virosa*); LACT_5CDS.CSA1 database and then aligning the paired end transcriptome sequences from Larissa WT to build up a reference sequence for Larissa WT. The combined databases contained 214,743 ESTs (122,982,522 bp); 144,711 (67.4 %) of the ESTs are covered by the Larissa WT 76 bp transcriptome sequence reads. A reference transcriptome sequence for Larissa WT was constructed comprising 54,411,874 bp in total; the paired end reads helped to resolve the issue of repeat segments in the transcriptome. The remainder of the sequences obtained did not align to any ESTs in the database.

The mutant transcriptome sequences were then aligned to the contigs created from the Larissa WT reference transcriptome sequences, 42.33 % of the paired end reads obtained from the LB Larissa line S2_BC2_164_E1 aligned successfully to at least one Larissa WT sequence. Upon aligning these sequences 1846 sequence differences (where the difference is covered by at least two reads in each direction) were identified between Larissa WT and LB line S2_BC2_164_E1, a mutation rate of approximately 1 per 12.5 kb (sequence differences were removed where base calling did not identify a clear nucleotide).

The sequence variation observed was ranked based on the number of reads/depth of coverage in which the sequence difference was observed in the WT and mutant sequences, this provided a level of confidence and indicated which differences were more likely to be real.

The eleven ESTs in the database which were used to obtain the target flowering gene homologues from lettuce (see chapter 3) were used to provide an indication of the depth of coverage that was observed from the Larissa WT paired end reads. This provided an indication of the level of coverage obtained from the transcriptome sequencing, for example the autonomous pathway genes are known to exhibit a wide range of expression in *Arabidopsis*; mRNA encoding each gene is detectable in almost all organs that were tested (Simpson, 2004). Table 5.6 contains a summary of this data, a depth of coverage of five reads per EST base was deemed acceptable. The EST representing *CO* contained the highest level of coverage; 92.9 % of the EST being covered by five reads or more, this could be because the other 16 *COL* genes are so highly conserved that reads from these related genes are increasing the number of reads aligning with this EST. The EST representing *FY* was not covered by the transcriptome sequencing, and the coverage of the EST representing *FPA* was relatively low. The Bio-Array Resource for Plant Functional Genomics, University of Toronto (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi), has shown that *AtFY* is highly expressed in the shoot apex and relatively highly in most other tissues, however no data is available for *AtFPA*, this could suggest it is expressed at a relatively low level. Overall 70 % of the 10 ESTs (excluding *FY*), where sequence was available, representing the target genes were covered by a depth of coverage of five reads or more.

Gene	EST	Reads per EST base	EST bases with a depth of coverage of five of more (%)
FT	CLSS2320.b1.O03.ab1	1-19	65.9
FLK	QGF19F22.yg.ab1	1-19	59.2
FLD	CLRX5844.b2.G21.ab1	1-19	43.5
LD	CLSM7821.b1_J11.ab1	1-21	82.2
FVE	QGC24A13.yg.ab1	1-37	51.5
FY	CLRX15054.b1_K19.ab1	No coverage	N/A
FCA	CLSM1171.b1_E05.ab1	1-65	91.3
FPA	CLVX2126.ab1_O12.ab1	1-3	0
CO	QGF10H03.yg.ab1	1-99	92.9
FKF1	QGG32A02.yg.ab1	1-55	72.6
CRY2	CLSM20253.b1_I24.ab1	1-32	76.1

Table 5.6 – Depth of coverage recorded using transcriptome sequencing for Larissa WT for each of the lettuce ESTs showing homology to *Arabidopsis* flowering time genes

It was decided that a potential mutation in a gene of interest would only be followed up if it had a depth of coverage of at least two in both Larissa WT and the mutant line.

The analysis of the transcriptome data for S2_BC2_164_E1 identified 18 potentially interesting changes between the mutant line and Larissa WT which all result in an amino acid change, see table 5.7. The 18 mutations were all annotated with GO terms relating to developmental processes which could potentially be affecting the LB phenotype observed in line S2_BC2_164_E1. Four of the sequence differences were of particular interest as they were identified in ESTs with homology to the *Arabidopsis* genes *COP INTERACTING PROTEIN 8 (CIP8), COL4, GI* and *GRP1*, a splice variant of the gene *GLYCINE-RICH RNA BINDING PROTEIN 7 (GRP7)*. The RING-H2 domain of *CIP8* interacts with the *CONSTITUTIVE PHOTMORPHOGENIC 1 (COP1)* which is a repressor of photomorphogenesis (Oyama *et al*., 1997;Torii *et al*., 1999). *COL4*, as already discussed is part of the *CO* and *COL* gene family, although it's function is yet to be characterised. It appears, from the work completed so far in this project that an *LsCOL4* gene homologue has been identified, see section 3.3.10, along with the expression pattern of this gene, see

figure 3.48. This means that the effect of any mutation on the expression of *LsCOL4* can easily be identified. *GI* is a single copy gene in *Arabidopsis* and is an important gene within the photoperiodic pathway, it acts upstream of *CO* and the *gi* mutant is late flowering. GI also acts as part of a feedback loop of the plant circadian clock (Park *et al*., 1999). *GRP7* has been shown to have a role as a flowering time gene as a potential member of the autonomous pathway (Streitner *et al*., 2008). The *GRP1* splice variant has been shown to bolt later than WT under both LDs and SDs; but the effect of the mutation can be overcome by a vernalisation treatment. The nucleotide change (adenine to cytosine) at position 438 of EST DW144163 results in the creation of a serine amino acid in line 164 which replaces the STOP codon encoded for in Larissa WT. The coverage of each of the four mutations was fairly low, as shown in table 5.7.

Each polymorphism results in an amino acid change, which may cause an altered function of the protein which results in the LB phenotype observed in line 164. Because of the low level of coverage for the polymorphisms identified in each of the four genes, PCR primers were designed to amplify small fragments of DNA surrounding the polymorphism site, see Appendix II for primer details. The PCR products amplified from Larissa WT, line S2_BC2_164_E1 (genomic DNA extracted from the pooled leaf samples used to extract the RNA from for the transcriptome sequencing) and line 164_B4 (a sister line of S1_BC1_164_E1, which after BC to WT and selfing was no longer LB and therefore should not contain the polymorphism if it is having a direct effect on the bolting phenotype), were then sequenced to verify the sequence differences suggested by the transcriptome sequence analysis.

EST	*At* Gene	EST locus	EST base	WT call	164 call	Depth of coverage WT	Depth of coverage 164	Amino acid change
DW142989	*H2A3*	142	G	G	A	331	168	Q-K
BQ861968	*RSP40*	377	T	T	A	9	8	H-G
DW082577	*SM312*	446	A	A	C	5	6	K-T
BQ861968	*RSP40*	376	C	C	A	5	8	P-G
BQ851710	*SUGT1*	251	T	T	A	11	4	Y-N
BQ875498	*RKF3*	536	T	T	G	33	4	Unsure
DY959953	*NPH3*	397	A	A	T	8	4	N-Y
DW131430	*RUXE*	410	T	T	C	3	6	In 3'UTR
DW144163	*GRP1*	438	A	A	C	4	3	STOP-S
DY961816	*AFC2*	840	T	T	C	3	3	S-P
DY971753	*AGO1*	736	T	T	G	3	6	Y-N
DY981409	*P2C25*	143	C	C	T	14	2	S-F
BQ985699	*SUVH5*	223	A	A	G	3	3	R-G
BQ987367	*WDR8*	436	T	T	A	9	3	F-I
BU013199	*UPL1*	245	T	T	G	42	3	L-R
DW139349	*CIP8*	554	A	A	G	8	2	H – R
DY961668	*COL4*	83	G	G	T	2	2	R – L
DY977769	*GI*	683	T	T	A	9	2	L – Q

Table 5.7 – Transcriptome sequencing analysis - Mutations identified in genes involved in developmental processes including flowering in the LB Larissa line 164 compared to WT. EST ID = GenBank number.

In each case the sequences obtained from the LB Larissa line S2_BC2_164_E1 matched that seen in Larissa WT and the S1_BC2_164_B4 line (which bolted as WT). These false positives suggested a problem with the data obtained and that a more stringent cut off was required, this could be achieved by requiring a greater depth of coverage of the polymorphic base in both transcriptome reads.

The remaining 14 changes observed between Larissa WT and LB line 164 were analysed by sequencing PCR products covering each individual SNP. On each occasion the polymorphism could not be verified, even the sequence difference observed in the EST DW142989 which was covered by 331 WT reads and 168 line 164 reads was identical to WT. Therefore it appears that there is an issue with some of the data obtained, either in terms of the reads generated from the Genome Analyser or from the bioinformatic analysis of the results. Work relating to these issues is ongoing, but it has been shown that sequence differences between Larissa

WT and LB Larissa 164 in sequences aligning to ESTs with non-flowering related functions have been verified where at least one of the sequences displays a depth of coverage of 6, 13, 42, 74, 133, 139 and 230 reads, it appears as though there is no consistent number of reads which will produce data which can confidently analysed.

A further 47 non-annotated (sequences which do not align to ESTs in the database and have no homology to other known sequences), sequence differences are currently being followed up, to investigate whether they are involved in the LB phenotype observed in line 164. These polymorphisms may be in novel genes involved in flowering which are found in lettuce, but not *Arabidopsis*. The depth of coverage for these 47 sequence differences ranged from 3 to 597.

15 of the sequence differences could not be verified in the mutant 164 DNA. A further six mutations were confirmed in line S2_BC2_164_E1, however the sequence difference was also present in the DNA extracted from the sister plant S1_BC2_164_B4 which was not LB. These six mutations are therefore not causing the LB phenotype observed.

The Genome Analyser data for LB Larissa line S2_BC2_307_B4 has recently been processed and 45 interesting sequence differences that have been identified in sequences which show homology to genes involved in developmental processes, these include an RNA binding protein (potentially of interest, as some of the autonomous pathway genes are known to interact with RNA), *CRY1*, *CRY2* and a *COL* gene sequence.

5.3.2.2 Analysis of LB Jessy lines containing the *Tnt1* retrotransposon element

5.3.2.2.1 PCR screen of target genes in interesting Jessy lettuce lines transformed with the tobacco *Tnt1* retrotransposon element

The eight flowering time genes targeted in this project were screened for mutations in the LB Jessy lines, using the method previously described to screen the LB Larissa and Saladin lines, see section 5.3.1.3. PCR reactions were carried out as described in table 5.4, the *Tnt1* element is 5334 bp in length, therefore the amplification of a larger than expected product or the absence of a PCR product would indicate the presence of a *Tnt1* element in the gene. The 17 PCR reactions covering the full length of the seven lettuce autonomous pathway genes, plus the three PCR reactions used in the TILLING experiments, see section 5.3.1.2 to cover *LsFT* were performed using gDNA extracted from the LB Jessy lines J_T3_Tnt1a_2 and J_T3_Tnt1a_5, along with a control; genomic DNA extracted from Jessy WT. Products of an expected size were amplified from both the LB Jessy lines. This result showed that the *Tnt1* element had not inserted itself into one of the eight lettuce flowering time gene homologues identified.

5.3.2.2.2 Screening the LB Jessy lines to identify the location of the tobacco *Tnt1* element using iPCR and Genome Walking

As the *Tnt1* retrotransposon element had not inserted itself into one of the eight lettuce flowering time gene homologues in the LB Jessy lines, alternative methods were used to try to identify the location of the elements. Initially an iPCR method described by Lucas *et al*., 1995 was tested; two primers Tnt77 and Tnt540 were used to amplify the flanking sequences of the *Tnt1* element. Genomic DNA extracted from plants from lines J_T3_ Tnt1a_2 and J_T3_Tnt1a_5 was digested with *EcoRI*

or *XbaI*, the product was then self ligated. PCR was performed on the purified product, using standard KOD conditions, the products amplified were then cloned. This method proved unsuccessful, the products generated were very weak and could not be further optimised, making it very difficult to successfully clone the products. Some of the plants tested produced no products at, suggesting the possible truncation of the *Tnt1* element in these plants as both primers were designed close to the ends of the 3' and 5' sequence.

PCR primers for genome walking were designed to both the 3' and 5' end of the *Tnt1* retrotransposon sequence. The sequence is very highly repetitive, which made primer design particularly difficult. The material screened was extracted from line J_T3_Tnt1a_5_6A_2, the latest S1_BC1 plant to bolt, see section 4.8.3.2. Mazier *et al*., (2007) showed that *L. sativa* cv. Mariska plants which had been BC once to WT plants contained between only nine and sixteen *Tnt1* inserts.

Standard genome walking PCR conditions were used, as described in section 2.2.7. The PCR reactions which worked most successfully after the optimisation of each reaction using primers designed to both the 3' (GW_RET_3_GSP1 and GW_RET_3_GSP2) and 5' (GW_RET_5_GSP1 and GW_RET_5_GSP2) ends of the retrotransposon are shown in figure 5.17.

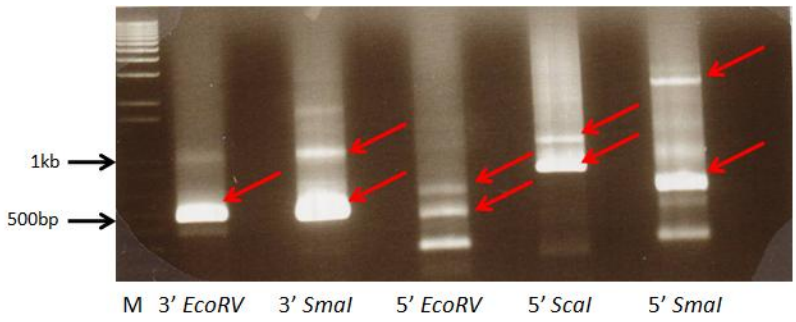


Figure 5.17 – PCR products amplified using genome walking from the *Tnt1* retrotransposon elements inserted within the LB Jessy line J_T3_Tnt1a_5_6A_2, M = 1 kb+ ladder.

Each of the nine products amplified, highlighted with red arrows, from the five genome walk libraries were cloned and sequenced. The 500 bp 3' *EcoRV*, 500 bp 3' *SmaI*, 500 bp 5' *EcoRV*, 650 bp 5' *EcoRV* and the 650 bp 5' *SmaI* products amplified were all too small to obtain enough sequence to identify the genomic region into which the retrotransposon had inserted itself. The remaining four fragments did produce a significant amount of sequence, each of which were BLASTed in NCBI and CGP2 at the nucleotide and amino acid level. The 1 kb 3' *SmaI*, 1 kb 5' *ScaI* and 2.5 kb 5' *SmaI* products all produced no matches in either database. The 850 bp *ScaI* did show homology to nitrate reductase (NADPH) genes in many plant species including grape, potato and rice, this gene has no characterised role in flowering, so the *Tnt1* insertion in this gene is unlikely to be causing the LB phenotype in line J_T3_Tnt1a_5_6A_2.

The PCR conditions used for the genome walk reactions were further optimised by diluting the concentration of the ligated product and first round PCR products used in the PCR reactions, this produced five more bands (the products are highlighted with red arrows in figure 5.18), from the 5' *DraI*, 3' *PvuII* and 3' *PvuII* libraries. These fragments were also cloned.

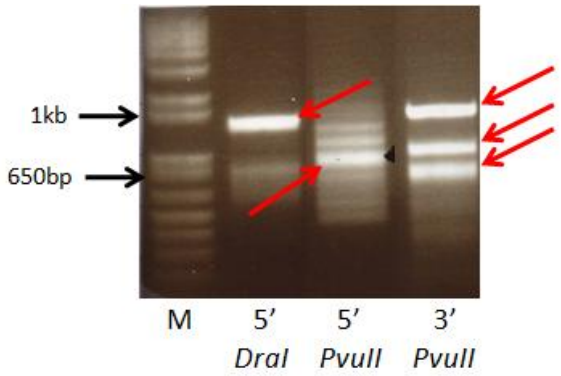


Figure 5.18 – PCR products amplified using genome walking from the *Tnt1* retrotransposon elements inserted within the LB Jessy line J_T3_Tnt1a_5_6A_2, M = 1 kb+ ladder.

Once again the fragments showed no homology to any lettuce ESTs in the CGP2 database, nor did they produce any hits in NCBI.

The retrotranspsoson strategy is potentially a good method in gene tagging experiments. The Jessy lines which still showed a delay in bolting after BC and selfing, should contain a relatively small number of elements, however identifying the location of the retrotransposon insertions within J_T3_Tnt1a_5_6A_2 proved difficult. An improvement on the genome walk method used would have been to also extract material from Jessy WT plants, Jessy plants which were LB in the T3 generation and Jessy plants which were no longer LB after BC to WT. This would mean that the bands amplified from the genome walk experiments could be compared with one another and unique bands could be followed up for further analysis. For example, bands amplified from the LB T3 plants and plants which were LB after BC and selfing could be prioritised as being potentially interesting. Bands common to all plants could be ignored as they would not be insertions affecting the bolting phenotype of the Jessy plants. A method for analysing retrotransposon insertions by Syed and Flavell (2006); Sequence-specific amplification polymorphisms (SSAPs), has been shown to be highly successful at identifying the location of retrotransposon elements, this is another method being investigated for future work.

5.3.3 Summary

The mutations causing the LB phenotypes observed in the Larissa, Saladin and Jessy populations could not be identified in this project. However this project has identified a number of successful methods to screen the LB lines for mutations. Methods to look for mutations within the genes targeted in this project proved successful, TILLING produced excellent reproducible data, but unfortunately it was

too expensive to follow up. The sequencing of PCR products, although time consuming, proved to be an efficient alternative. Transcriptome sequencing to identify all the mutations existing in LB lines compared to WT, is potentially the most robust method tested. Work needs to be done to verify the data obtained and to identify confidence levels which can be applied to the data to ensure that mutations identified are 'real', but in the long term this approach appears to be the best option.

CHAPTER 6

Discussion

6.1 Identification of flowering time gene homologues in lettuce suggests the *Arabidopsis* flowering network may be conserved between the two species

Both *Arabidopsis* and lettuce are facultative LD plants, flowering earlier under a photoperiod of 16h. It has been suggested from work in several plant species that the genes involved in the floral transition are conserved across different species, for example research in the SD plant rice has shown that a number of key genes involved in this process perform the same function as in *Arabidopsis* (Kojima *et al.*, 2002; Izawa *et al*., 2003; Hayama *et al*., 2003). However, to date very little of the research performed in other species has led to any beneficial improvements being made in crop species.

One of the aims of this project was to identify lettuce homologues of key genes involved in the initiation of flowering in *Arabidopsis*, and where possible to show that the homologues identified were performing the same function. An initial list of 20 *Arabidopsis* target genes was constructed and using the CGP database ESTs with a high level of homology to each were identified. The initial list of 20 genes was rationalised to 12 which fitted the criteria described in section 1.7. These included the seven genes which make up autonomous pathway in *Arabidopsis*. The commercial benefits of mutants in these genes are two-fold; growers benefit due to a delay in bolting, but breeders can use a vernalisation treatment to overcome the effect of the mutation to restore the bolting phenotype, meaning that the timing of seed production is not affected.

The ESTs were used to probe a lettuce BAC library and clones containing the flowering time gene homologues were obtained. PCR based methods were utilised

to obtain full length gene sequences of eight of the genes targeted, including the seven autonomous pathway genes, (later work by Dr Andrea Massiah resulted in the isolation of a full length sequence of a ninth gene). The lettuce homologues displayed 43-79 % identity with their *Arabidopsis* counterpart at the amino acid level. Each of the lettuce homologues identified also displayed a high level of homology to the well conserved functional domains characterising specific genes in *Arabidopsis,* such as the three KH RNA binding domains in *FLK*. This increased the confidence that the lettuce homologues identified were more similar to the *Arabidopsis* genes targeted than to other related genes, which in some cases were other members of a gene family, even though gene family members often share those conserved domains.

The *LsFT* gene isolated was able to functionally complement the *Arabidopsis ft-1* mutant and restore the late flowering phenotype of the mutant to that seen in L*er* WT plants. Its expression pattern in lettuce was also similar to that of *FT* in *Arabidopsis*. This suggests that the gene identified is a lettuce *FT* orthologue. However *LsFT* may not be the only functional orthologue of *AtFT* in lettuce; it is possible that more than one lettuce *AtFT* orthologue exists, this does appear unlikely as only one lettuce EST showed any homology to *AtFT* from over 228,000 ESTs obtained from five lettuce species. Furthermore the EST only produced one hit on the lettuce BAC library which was estimated to have a two to three fold coverage of the lettuce genome. *FT* is expressed at a low level in *Arabidopsis,* so this may account for why only one EST with any homology was identified in the EST database. Also of interest is the absence of any ESTs with homology to the other *AtFT* family genes, for example *AtTSF* shows 81.1 % homology with *AtFT,* so it would have been expected that a lettuce *TSF* would have been identified when screening the EST library using *AtFT*

to BLAST against the database. It would also be expected that using the *LsFT* EST to probe the lettuce BAC library would have identified genes similar to *FT* present in the BAC library. The analysis of a Southern blot of digested genomic lettuce DNA using radioactively labeled probes of different fragments of *LsFT* would have given an indication of the copy number of the gene in lettuce, but unfortunately this data could not be obtained during this project.

If we assume that the *FT* sequence isolated from lettuce is the main *Arabidopsis* orthologue then comparisons between the way in which flowering is initiated in *Arabidopsis* and lettuce can be made. It is known that the rhythm of *AtCO* expression peaks at the end of a LD and this provides the basis for the regulation of the photoperiodic pathway by day length (Suàrez-Lòpez *et al.*, 2001). *AtCO* plays an important role in this pathway by mediating between the circadian clock and the floral integrators. In *Arabidopsis, CO* expression coincides with light at the end of a LD resulting in the accumulation of CO protein in the nucleus, which activates the expression of *FT*, which in turn promotes flowering. The real time data collected for the *LsFT* orthologue and an *LsCOL* homologue (which is phylogenetically more similar to *AtCOL4*, see section 3.3.10.2), shown in figure 3.47 show a peak in expression at around ZT14, towards the end of a LD, see figure 6.1. As the function of *LsCOL* has not been investigated in this project it is unknown whether it has an effect on flowering and bolting; complementation of the *Arabidopsis co-2* mutant with *LsCOL* would be one approach to investigate this. *AtCOL4* has not been characterised, but microarray data shows that its expression increases prior to bolting and decreases upon floral initiation (Hruz *et al*., 2008). This suggests a possible role for *AtCOL4* in the control of flowering in *Arabidopsis*. It could be suggested that the overlapping expression of the *LsCOL* homologue and the *LsFT* orthologue may

indicate a functional interaction that would be worth further investigation. Interestingly a *CO* orthologue in sugar beet has been shown to be closely related to *AtCOL1* and *AtCOL2* both through phylogenetic analysis and in the time of day at which it is expressed; at dawn is able to functionally complement the *Arabidopsis co-2* mutant (Chia *et al*., 2008). Therefore it seems that *CO-like* genes are able to play a role in the induction of flowering in some species.

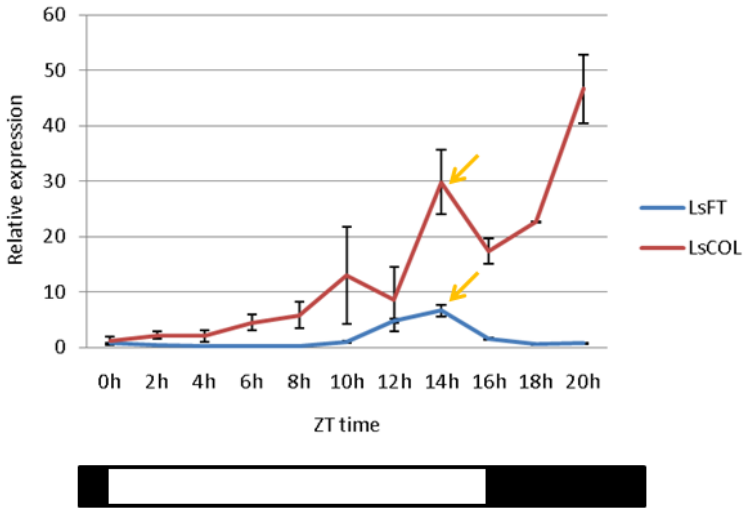


Figure 6.1 – Comparison of *LsFT* and *LsCOL* expression over a 24h period. The peak in *LsFT* expression coincides with the increase in *LsCOL* expression, highlighted with orange arrows.

Figure 6.2 shows a simplified diagram of the main genes, pathways and interactions that lead to the induction of flowering in *Arabidopsis*. Work from this project has identified a functional *LsFT* which has been characterised, it has also identified a *LsCOL* gene homologue with an expression profile similar to *AtCO*, four candidate *LsFLC* genes and homologues of seven autonomous pathway genes (including a functional *LsFLK* orthologue), along with *LsCRY2* and *LsFKF1*, (the function of which are currently being analysed) candidate gene homologues. As well as this ESTs with high homology to *PHYB* and perhaps more significantly *GI* exist in the CGP database, three ESTs; two from *L. sativa* and one from *L. serriola* make up a

contig which covers the 1173 amino acid *AtGI* from amino acid 468-1043 and show 71 % identity at the amino acid level.

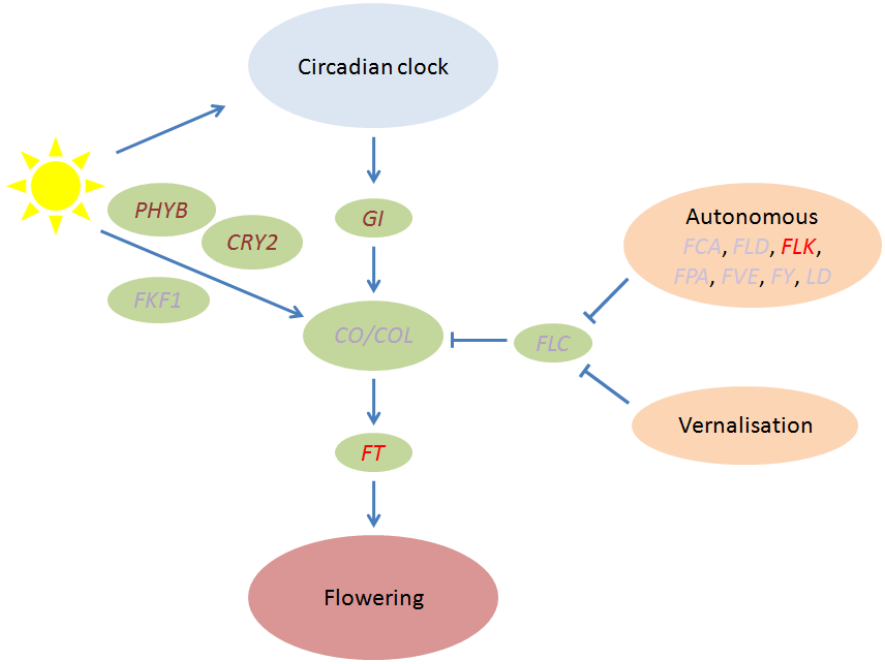


Figure 6.2 – Simplified version of the pathways and genes involved in flowering in *Arabidopsis*, lettuce orthologues (highlighted in red), homologues (highlighted in lilac) or ESTs (highlighted in magenta) have been identified with high levels of homology to each of these genes. This suggests conservation of the flowering regulatory pathway between the species.

This work, along with an additional 74 lettuce ESTs which show homology to other *Arabidopsis* flowering time genes identified by our collaborators in Professor Michelmore's group suggests that a high level of conservation exists in the pathways controlling flowering in lettuce and *Arabidopsis*. Of the lettuce autonomous pathway gene homologues identified, complementation work was only carried out with *LsFLK* and *LsFLD*. When over expressed in both Col-0 WT and *flk-4* mutant plants the *LsFLK* homologue produced plants which flowered significantly earlier than non transformed Col-0 and *flk-4* plants. The transgene appeared to be segregating in a 3:1 ratio of plants flowering earlier than WT compared to plants flowering as WT in the T2 Col-0:*LsFLK* generation. Preliminary data has been collected from the T2 *flk-4*:*LsFLK* generation which suggests the *LsFLK* gene has recovered the LB phenotype of the *flk-4* mutant, this data suggests a *FLK* gene orthologue has been isolated from

lettuce. The *LsFLD* homologue did not functionally complement the *Arabidopsis fld-3* mutant. This could be because of the difference in species; interestingly successful complementation of the *Arabidopsis fld-3* mutant by *At*FLD has not been reported. The LsFLD homologue is 69.2 % identical to *At*FLD and is the top hit in a BLAST search; it is possible that although it is a gene homologue of *AtFLD* it doesn't perform the same function in lettuce, or that the homologue identified is not the true *Arabidopsis FLD* orthologue. It is interesting that the *LsFLD* homologue is encoded by one large exon, whereas the *AtFLD* gene is encoded by two, the two functional orthologues of *AtFT* and *AtFLK* identified in lettuce were both encoded by an identical number of exons. From this it could be inferred that the *LsFLD* homologue is not the functional orthologue of *AtFLD*. The lettuce gene homologues of the other autonomous pathway genes require further work to prove their roles in flowering in lettuce. Complementation studies using the lettuce homologues to restore flowering time in the corresponding *Arabidopsis* mutants could be carried out as could gene knock out experiments which target the gene homologues in lettuce. Lettuce knock-out lines generated by RNA interference (RNAi), (Hammond *et al*., 2000; Wroblewski *et al*., 2007) or by using virus induced gene silencing (Lu *et al*., 2003) could be screened for a delayed bolting time and could be complemented by *Agrobacterium* transformation with the gene of interest (Horsch *et al*., 1985; Curtis *et al*., 1994). The creation of, for example, double mutants involving *FLD* with either *FCA* or *FPA* would be interesting, given that recent research in *Arabidopsis* has shown that FCA and FPA both potentially act through FLD (Bäurle and Dean, 2008).

6.2 LB lettuce lines have been identified in three commercial lettuce varieties; five homozygous lines in the cv. Larissa have a robust LB phenotype

One of the aims of the project was to identify LB lettuce lines from a number of different lettuce populations, including lines containing induced mutations and those containing natural allelic variation. What constitutes LB varies depending on the grower/breeder. LB is a desirable trait to all growers especially if the plants can hold their mature phenotype for a number of days in the field; depending on the size of the harvesting operation, anything between two and seven days is desirable; hence this project has tried to use at least a seven day cut off to distinguish LB lines from WT. A couple of issues that complicated the screening for LB lines was the timing of the glasshouse experiments, and the large differences in the number of days to bolt observed for the plants when they were grown at different times of the year or in the glasshouse compared to the field. To maximise the output of this project lettuce plants were grown throughout the year; where necessary artificial lighting was used to supplement the natural daylight to provide LB conditions in the glasshouse. Each line with a potentially LB phenotype was confirmed in the field under a natural LD photoperiod, as it is impossible to replicate field conditions within a glasshouse environment.

The most extensive research focused on the EMS mutagenised populations in the Larissa and Saladin background. Because of its relatively quick life cycle, the ability of the plant to bolt under artificial supplementary lighting and the small size of the population, more work was done with the Larissa lines. A very small percentage of the original Larissa plants treated with EMS germinated and went on to set seed because the seeds had been treated with a range of EMS concentrations; M2/M3 seed from only 62 lines was available for this project. Three different LB Larissa lines

(lines 2, 164 and 307) were identified that had a LB phenotype irrespective of whether grown under natural or artificial LD photoperiod in the glasshouse. After two rounds of BC to WT and a self fertilisation three lines originating from line 2 (2_C1_2, 2_D3_3, 2_G2_1) and single lines originating from lines 164 and 307 (164_E3_1 and 307_B4_4) produced plants which all retained the LB phenotype. The inheritance of the mutant phenotype observed in each of these homozygous LB mutant lines suggested that they were all caused by dominant mutations. As the Larissa population was used as a test population for the larger Saladin EMS mutagenised populations, only lines with an obvious LB phenotype were chosen to take forward, therefore several other LB lines, which may have been caused by recessive mutations, were not selected to be followed up in this project.

The LB phenotype of each of the selected lines was robust under commercial growing conditions in the field; plants representing each line grown during the late spring through to the middle of summer, in two different locations, all bolted later than WT. The plants also all grew at a similar rate to WT with no obvious phenotypic differences. The LB phenotype was also robust under increased temperatures, although line 164_E1_3 begins to lose its LB phenotype at 25 °C. This line also showed a less robust LB phenotype in the summer field trials where the average maximum daily temperature in the two trials was 23 °C and 24.1 °C in Spalding and Wellesbourne respectively. Line 164_E3_1 does show a response to a vernalisation treatment; plants bolt significantly later than WT when grown with no treatment in the glasshouse, however when subjected to a vernalisation treatment of four weeks the plants bolted at a similar time to WT, this suggests that the mutation causing the LB phenotype lies within one of the genes in the autonomous pathway.

Line 2_C1_2 displays an interesting bolting phenotype under an increased temperature; in the temperature experiments performed at a constant temperature of 25 °C and in the second field trial held at Wellesbourne with an average temperature of 24.1 °C (which included 12 days with an average maximum daily temperature of over 25 °C). The plants bolted later than all the other LB lines in the second Wellesbourne trial, whereas under lower temperatures in the first Wellesbourne trial and the Spalding trial the line was the second earliest of the five LB lines to bolt. In the constant temperature experiments carried out in the controlled environment cabinets line 2_C1_2 was the only line not to bolt significantly later than WT at 23°C, bolting on average just 4.25 days later, but at 25 °C the line bolted 9.5 days later than WT. Lines 2_D3_3 and 2_G2_1 which originated from the same line 2 as 2_C1_2 did not show this response under high temperatures, meaning the mutation causing this phenotype in line 2_C1_2 is likely to have been segregated away in the other two lines by BC. It would be interesting to grow line 2_ C1_2 in even higher temperatures in the field to see if the LB phenotype holds. This is an important trait in any new LB lines generated with the recent trend for warmer summers caused by climate change. Line 307_B4_4 displayed an extremely LB phenotype under all of the conditions in which it was grown, this line is perhaps the most obvious line to follow up in commercial trials as the phenotype is consistent even under extreme temperatures.

Less work has been done on the Saladin EMS mutagenised populations as they appear to have a light quality requirement for bolting, which means that they will not bolt under the high pressure sodium lamps used in the glasshouse to supplement the natural SD photoperiods of the winter months. These plants have been through either one or two rounds of BC and a self fertilisation, the self seed has been scored

just once in relatively small numbers, but 52 LB individuals from 21 BC1_S1 and BC2_S1 lines have been identified. Further work on the Saladin LB lines will include the collection of BC2_S2/BC1_S2 seed from the LB BC2_S1/BC1_S1 individuals identified and growing them to identify homozygous mutant lines. Because M1 data is available for the main population the genetics causing the LB phenotype can be assessed more easily than for the Larissa population where this data was not available.

Two interesting lines were identified from the Jessy lines which had been transformed with the *Tnt1* retrotransposon element from tobacco; both lines originated from the same T1 plant, of which seed was not available to screen. Both lines were LB without a vernalisation treatment; the bolting phenotype was restored to WT with a vernalisation treatment, suggesting that the lines contained a mutation within an autonomous pathway gene. These lines should have a relatively low number of insertions after BC (this has been shown to be as low as 9-16 in similar experiments, Mazier *et al*., 2007), so it can be assumed that the mutation causing the LB phenotype is the same in both lines. The mutation appears to be dominant, but homozygous in line J_T3_Tnt1_5 and heterozygous in line J_T3_Tnt1a_2.

Both mutations causing the LB phenotype in the Larissa line 164_E3_1 and the two Jessy lines which can be restored to WT with a vernalisation treatment, suggesting that they are mutations in autonomous pathway genes, appear to be dominant. Interestingly autonomous pathway mutants in *Arabidopsis* are caused by recessive mutations.

6.3 The LB phenotype of the lettuce lines identified is not caused by sequence variation in the *Arabidopsis* flowering gene homologues isolated

Material collected from each of the three LB M2/M3 Larissa lines 2, 164 and 307, ten of the most interesting LB M2/M4 Saladin lines 38, 64, 75, 124, 131, 335, 372, 808, 843 and 844 and the two LB Jessy lines J_T3_Tnt1a_2 and J_T3_Tnt1a_5 were screened by sequencing PCR fragments amplified to cover each of the eight lettuce flowering time gene homologues identified. Sequence variation was observed in only LB Larissa line 307; two mutations were identified in two different genes in material collected from the original M2 plant screened. The mutations were both identified within non-coding intronic sequence of *LsFVE* and *LsLD*; both were guanine to adenine substitutions, as expected with EMS mutagenesis (Krieg, 1963). However neither mutation co-segregated with the LB phenotype throughout the population in the subsequent generations of line 307, meaning these mutations were not causing the LB phenotype observed.

Having only isolated and screened for mutations in eight lettuce genes, the targeted approach of this project was always going to be ambitious. The EMS load for the Larissa population was likely to be at its maximum level; of 1000 M1 seed treated only 62 lines produced M2/M3 seed to be screened, and the phenotypes of some of these lines were greatly affected, see figure 4.11. The mutation rate for the LB Larissa line 164 after two rounds of BC, based on the transcriptome work section 5.3.2.1, was only one mutation every 12.5 kb. The mutation rate is likely to be even less in the Saladin population, as the seed was treated with a lower concentration of EMS for a shorter period of time. Interestingly though Larissa line 164 and the two Jessy lines J_T3_Tnt1a_2 and J_T3_Tnt1a_5 show the characteristics of *Arabidopsis* autonomous pathway mutants. The fact that no autonomous pathway mutation is

identified in any of the three lines provokes some interesting points. Firstly, are the autonomous pathway gene homologues that have been identified in this project the true gene orthologues of the corresponding gene in *Arabidopsis*? The lack of complementation of the *fld-3 Arabidopsis* mutant with the putative lettuce *FLD* would suggest not. However the transcriptome sequencing work in line 164 did not identify sequence variation in any region with homology to any of the seven autonomous pathway genes either. Alternatively it could be that unidentified components of the autonomous pathway exist which have not been characterised in *Arabidopsis*, for example the transcriptome sequence did show a mutation in a gene sequence with homology to *GRP1*, a splice variant of *GRP7* in line 164. It was shown that this mutation was not causing the LB phenotype, but on consulting the literature the authors (Streitner *et al*., 2008) suggested a potential role for the gene in the autonomous pathway; how many other uncharacterised genes potentially act in the same manner? The other possibility is that one or more of lettuce autonomous pathway genes identified may play a different role in lettuce to the role they play in *Arabidopsis*; this may explain the failure of *LsFLD* to complement the *fld-3* mutant. Little work regarding the function of the autonomous pathway genes in species other than *Arabidopsis* has been produced. The autonomous pathway genes do have other roles in *Arabidopsis*, they are involved in growth and development (Veley and Michaels, 2008), and it is possible that their role in lettuce is limited to this and not flowering.

6.4 New generation sequencing methods; the best route to identify potential alleles causing the LB phenotype in lettuce lines

The targeted gene approach, although successful in identifying potential gene homologues of flowering genes and suggesting some conservation between these genes in *Arabidopsis* and lettuce, did not identify any sequence variation within the LB mutant lines compared to WT. In parallel to this approach a more comprehensive method to identify sequence variation between the LB mutant lines and WT was sought. Initially the lettuce Affymetrix high density GeneChip® tiling array was considered, the rationale being that cDNA synthesised from the LB lines could be hybridised to the chip and mismatches between the oligonucleotides on the chip and the LB line could be tracked back to the EST sequence. Further analysis of the EST could be made to identify the *Arabidopsis* gene to which it showed highest homology. Any differences within flowering time genes could be investigated as a priority. However, the lettuce tiling array contains only 500 bp of each of the 35,000 lettuce ESTs making up the CGP database and so not the full length gene sequence, the other consideration is that the ESTs do not cover the promoter region of the genes and so mutations in the promoter which may affect expression levels/pattern would not be detected. This limits the amount of data that can be obtained. A more comprehensive way of identifying the underlying mutation affecting the LB phenotypes observed was to look at the sequence of all the expressed genes in the LB lines and compare them directly to the sequence of genes expressed in WT, using this technique all sequence variation occurring throughout the entire transcriptome can be directly compared. Again though, mutations in promoter sequences would not be detected. The Illumina genome analyser technology was used to obtain the transcriptome data. Because the lettuce genome is yet to be fully sequenced, the

lettuce EST database was used to create contigs from Larissa WT transcriptome data which would be used as the reference sequence. The LB lines were then directly compared with the WT reference sequence to identify sequence variation. The first line sequenced was the Larissa line 164_E1_3, whose LB phenotype could be restored with a vernalisation treatment, a phenotype of particular interest for the purposes of this work. Sequence variation was identified in four genes with significant roles in flowering in *Arabidopsis*; *COL4*, *CIP8, GI* and *GRP1*, however these 'mutations' could not verified by PCR and sequencing. Further analysis of the transcriptome sequences showed that the data, (irrespective of the depth of coverage/number of reads), was difficult to verify; of 46 differences observed between line 164_E3_1 and WT only 13 (28.3 %) have been verified by PCR and sequencing, this has subsequently been attributed to possible issues with the bioinformatic analysis of the data obtained, the extent of this issue is currently being assessed and re-evaluated. Potentially of most interest are some of the sequence differences observed in genes which do not identify an *Arabidopsis* equivalent when following a BLAST search, these changes may be in novel genes, which do not match any known gene sequences.

The genome analyser method does seem to be the most effective method to obtain the transcriptome data that is required for such a project; while the data analysis is causing some problems, this project was one of the first of its kind to be run at Warwick HRI and improvements are continually being made to the data analysis side of the process. The transcriptome sequencing method has the potential to be a very powerful tool in work of this kind if it can be improved sufficiently to provide consistent high quality sequence data. With the running costs for such technologies

dropping all the time, the method becomes more affordable and hence more lines can be analysed this way; including the LB Saladin, LB wild lettuce and Jessy lines.

It would also be interesting to look at the variation between the three LB Larissa lines 2_C1_2, 2_D3_3 and 2_G2_1 which all originated from the same M2 plant using the genome analyser technology. It can be assumed that the mutation causing LB in each line is identical, but that BC must have reduced the number of background mutations present in each line. It can be further assumed that each line has lost different background mutations, therefore comparing the mutations which are common to all three lines would cut down the amount of follow up work required; one of these mutations must be causing the LB phenotype. Furthermore, mutations present in line 2_C1_2 but not in lines 2_D3_3 or 2_G2_1 could account for the phenotype of this line observed under elevated temperatures.

6.5 Further Research

This project represents the first of its kind in lettuce and as such has produced lots of scope for future research.

As already mentioned in section 6.3, the targeted candidate gene approach relies upon the fact that the true lettuce orthologues of *Arabidopsis* flowering time genes will have been isolated. Therefore studies to investigate the function of the homologues identified so far should be followed up. If the homologues are not orthologues of the *Arabidopsis* genes then other ESTs and BACs identified with high homology could be followed up. Had time allowed it would have been interesting to look at the copy number of the lettuce gene homologues identified. This would have provided further evidence regarding the likelihood of the homologue being a potential orthologue, it would also provide some idea on the level of repetition within

the lettuce genome. The functional orthologues of lettuce flowering time genes would benefit breeding programs and provide a basis for future scientific research. The four lettuce *FLC* gene homologues identified should be followed up, the expression of each of the four genes in Larissa WT and the LB Larissa lines using a developmental time course of material was analysed using Real Time PCR see section 5.3.1.4. It would be interesting to look at the expression of each of the four genes with and without a vernalisation treatment. For any of the four genes to be considered as an *AtFLC* homologue a decrease in the level of expression would be expected in plants subjected to a vernalisation treatment.

The plant side of the research was restricted by the number of plants that could be screened at each stage, therefore some lines which appeared to be LB in initial screens could not be followed up due to the lack of glasshouse space and/or personnel available to maintain, score and BC the plants. It would have been interesting to have been able to screen more of the BC lines produced, and to have been able to screen more plants per line.

As already mentioned in section 6.2, the LB Larissa lines that have been identified require further assessment, but there is plenty of potential for these lines to be used in a number of ways. Further BC of these lines to WT would result in the removal of even more background mutations from the lines, making the identification of the polymorphism causing the phenotype in each line much easier, especially if using transcriptome sequencing as the screening method.

BC was extremely difficult to perform when attempting to use the mutant plant as the pollen donor, as the mutant plants, especially as M2/M3 and BC1 plants produced very little pollen and opened later in the day compared to WT, meaning the WT plants had developed too far for cross pollination to be successful, see figure 1.1.

However artificial methods to ensure the plants development is slowed could be used in future, for example, WT plants could be placed in the dark until they were at the appropriate stage for crossing. Reciprocal BC events would increase the chance of a successful BC event and would provide evidence that the BC technique was working. There is also potential to cross the LB lines with each other, this may produce progeny which bolt even later, and will produce another interesting resource in the form of a mapping population which can be used to identify loci involved in bolting time. Limited amounts of work, not reported in this project have been carried out to identify loci involved in flowering and bolting within the F5 and F7 generations of a *L. sativa* cv. Saladin *x L. sativa* cv. Iceberg mapping population. The population, created by Professor Pink's group at WHRI is currently being aligned to the *L. sativa* cv. Salinas x *L. serriola* map created by Professor Michelmore's group, at UC Davis, which is in nine linkage groups representing the nine pairs of lettuce chromosomes. This map is saturated with markers and rough QTLs have been established on both maps for bolting and flowering, Professor Michelmore's group have begun to map some of the lettuce flowering time gene homologues to these regions. QTL analysis was considered as an approach to identify alleles affecting bolting time in this project, however due to time restrictions only one approach could be undertaken in detail and a conscious decision was made to continue with the applied, targeted approach described in this thesis. A mapping project would also have meant replicating work already underway at UC Davis, the collaboration set up with Professor Michelmore means that the gene homologues identified in this project will be mapped, any corresponding to loci identified for flowering or bolting would provide more confidence that these genes are orthologues of *Arabidopsis* flowering time genes. As previously mentioned on top of the 20 ESTs identified for this

project with homology to *Arabidopsis* flowering time genes, Professor Michelmore's group has identified a further 74 ESTs with homology to *Arabidopsis* flowering time genes. It will be interesting to see if the gene homologues map to the flowering/bolting QTL loci already identified, however just as interesting will be the loci to which no genes map, these regions may contain novel genes with a role in flowering/bolting in lettuce.

Thought needs to be put into the future work on the LB lines Jessy lines identified, which are GM. These lines are of particular interest because their LB phenotype can be recovered by a vernalisation treatment. The mutation causing this phenotype could not be identified by screening the autonomous pathway lettuce gene homologues, methods to locate the *Tnt1* retrotransposon causing the LB phenotype also proved to be unsuccessful. However a method such as SSAPs (Syed and Flavell, 2006) could identify the genes into which the elements are inserted. Because of the agronomic implications of this project the use of GM plants was kept to a minimum. The general public's perception of GM foodstuffs is still a controversial one, and as such the amount of work undertaken in this area in this project reflects those views. However analysis of GM crops is still relevant and does need further investigation especially with climate change and the increase in population forecast for the next 50 years, which will inevitably put pressure on food production. The potential transgene flow from lettuce crops to weeds is the focus of a research initiative being investigated in Europe under the name of ANGEL looking at the population-ecological consequences in the context of GM-crop biosafety (www.plant.wageningen-ur.nl/projects/angel/). A potential link with this initiative may result in advice suggesting the route to take with these lines in terms of providing an agronomic benefit.

More work can be carried out using the LB wild lettuce lines identified, there is clear sequence variation between the wild lettuce lines and the cultivated lettuce cv. Larissa. This can be done by Ecotilling (Comai *et al*., 2004) or by using the genome analyser to identify differences between the LB and EB lines within an individual species, these differences can be analysed in more detail to discover if they are causing the LB phenotype. In parallel to this Rijk Zwaan® are interested in integrating the LB lines into their breeding programs.

6.6 Assessment of this research project with regard to the lettuce industry

Initial results obtained from the first round screen for bolting time in both the EMS Larissa and Saladin populations resulted in interest from the plant breeder Rijk Zwaan®, a collaboration was agreed and some of the LB lines along with the target gene sequences have been made available to them for further assessment. Their intention is to include the interesting alleles in their breeding programs. Rijk Zwaan® are also interested in taking forward the lettuce wild diversity set work started in this project; the LB lines identified proved difficult to BC to commercial lettuce lines, however they have the skill, expertise and personnel to successfully set up such a breeding program.

In addition, academic links have been made with UC Davis. The CGP, a collaboration, run by Professor Michelmore's group is at the forefront of research in lettuce. This project has benefitted through access to genomic tools, such as a lettuce BAC library and plant materials, such as wild lettuce accessions. This collaboration has resulted in *FT* data currently being collated for publication. This research project, with the backing of Professor Michelmore and Rijk Zwaan® resulted in the

award of a five year BBSRC grant (2008) to continue developing the research described in this thesis.

Future genetic experiments in lettuce will benefit from the full length lettuce gene sequence which is one of the objectives of the CGP. It will also benefit from sequence information generated from the model plant for the *Compositae*, *Leontodon taraxacoides*, which has a genome size of only approximately twice the size of *Arabidopsis* (Michelmore, 2007).

Now that Larissa lines with a robust reproducible LB phenotype have been identified and field experiments have shown that the lines develop at a uniform rate which is critical to enable a successful harvest with fewer crop losses, it is important to look to the next stage of development of the cultivar. A LB phenotype was the key to this project and this has been achieved. Further experiments will need to evaluate the disease resistance of any new line; it could be that the EMS treatment has had a detrimental effect on a gene involved with resistance to one of the many diseases/viruses affecting lettuce. Another important trait to test is that of taste; again there is potential that the EMS treatment may have affected this important trait; a potential collaboration to test this is currently being agreed with Dr. Carol Wagstaff, University of Reading. There is a line of thought that the bitter tasting secondary compounds produced by the lettuce plant when flowering is initiated, may actually be present in the plant throughout its development, but the reason that the plants taste bitter when flowering is initiated is due to the cessation of the plant's sugar production, this potential collaboration would also investigate this hypothesis.

The LB lines generated in this project may be used as a tool in developing new varieties as opposed to ultimately being marketed as a new variant of an already available cultivar. If the allele involved in causing the LB phenotype can be

identified, it becomes a powerful new breeding line. The allele can be used as a marker for LB; the line identified containing the LB allele can be crossed to popular lettuce cultivars; the plants can be genotyped early in their development, the presence of the LB allele means these plants should LB, thus improving popular commercially available cultivars. An uncharacterised EB gene locus *Ef-1Ef-2* has been used by breeders in this way to speed up seed production (Ryder 1983; Ryder 1985).

The EMS approaches taken in this project have led to another potential collaboration with Elsoms Seeds Ltd, Spalding. They are interested in applying these methods and techniques to create a mutagenised population in rocket, with the aim of identifying alleles which delay bolting.

6.7 Conclusions

The project objectives were set out in Chapter 1, Section 1.2.2. A brief summary of the conclusions of the project in terms of the objectives is below

- *To identify and isolate lettuce orthologues of targeted Arabidopsis flowering time genes.* Eight lettuce homologues of *Arabidopsis* flowering time genes have been identified, *LsFT* and *LsFLK* have been confirmed as a functional orthologues of *AtFT and AtFLK*.
- *To identify LB commercial lettuce lines from populations containing induced mutations created by EMS mutagenesis and retrotransposon insertions.* Five lines, originating from a *L. sativa* cv. Larissa EMS population have been identified with a uniform and reproducible LB phenotype when grown in the field under commercial conditions. A further 52 lines originating from 21 M1 plants from a *L. sativa* cv. Saladin EMS population and two lines from a *L. sativa*

cv. Jessy population containing a *Tnt1* tobacco retrotransposon element are currently being further characterised.

- *To screen homozygous LB mutant lines for mutations within the lettuce flowering genes identified; if variation is not apparent within these genes then identification of mutations in non-targeted and novel flowering time lettuce genes will be investigated.* Two mutations within intronic sequence of *LsFVE* and *LsLD* were identified in the M2 generation of the LB Larissa line 307; these mutations were not present in subsequent LB BC and selfed generations and therefore were not causing the bolting phenotype. TILLING for mutations in *LsFT* identified two SNPs in an EB Larissa line 199 and a LB Larissa line 185, neither mutation was shown to affect bolting time. No mutations were identified in the other targeted genes in the LB Larissa, Saladin and Jessy lines. Transcriptome sequencing has been performed on LB Larissa lines 164 and 307; the analysis of this data is ongoing.
- *To search for naturally occurring allelic variation in lettuce target genes that affect bolting time in a diversity set of wild lettuce varieties.* LB lines from three wild lettuce species have been identified and preliminary work looking at sequence variation between LB and EB lines within each species has begun. Allelic variation causing the LB phenotype has not been identified at this stage and crossing these lines to cultivated WT lines has proved difficult. However, Rijk Zwaan® have shown an interest in these lines with a view to incorporating them into their breeding programs
- *To assess the utility of these new alleles for increased holding ability with lettuce breeders and growers.* The LB Larissa lines have been grown under commercial farming conditions in the field and shown to retain their LB phenotype. This has

led to interest from growers (J.E.Piccaver Co.) and breeders (Rijk Zwaan®). Further assessment of these lines is underway. This project has benefitted the four main targets for lettuce breeding, outlined in section 1.3.2; horticultural improvement, resistance to stress problems and uniformity of maturity including the adaptation to specific environments (Ryder, 1999; Michelmore, 2007).

References

Ahmad, M., Grancher, N., Heil, M., Black, R.C., Giovani, B., Galland, P. and Lardemer, D. (2002). Action spectrum for cryptochrome-dependent hypocotyl growth inhibition in Arabidopsis. *Plant Physiology*, **129**(2), 774-785.

Ahn, J.H., Miller, D., Winter, V.J., Banfield, M.J., Jeong, H.L., So Y-Y., Henz, S.R., Brady, R.L. and Weigel, D. (2006). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO Journal*, **25**, 605-614.

Almada, R., Cabrera, N., Casaretto, J.A., Ruiz-Lara, S. and Villanueva, E.G. (2009). VvCO and VvCOL1, two CONSTANS homologous genes, are regulated during flower induction and dormancyin grapevine buds *Plant Cell Reports*, **28**,1193–1203.

Argyris, J., Truco, M.J., Ochoa, O., Knapp, S.J., Still, D.W., Lenssen, G.M., Schut, J.W., Michelmore, R.W. and Bradford, K.J. (2005). Quantitative trait loci associated with seed and seedling traits in Lactuca. *Theoretical and Applied Genetics*, **111**,1365–1376.

Aukerman, M.J., Lee, I., Weigel, D. and Amasino, R.M. (1999). The *Arabidopsis* gene *LUMINIDEPENDENS* is expressed primarily in regions of cell proliferation and encodes a nuclear protein that regulates *LEAFY*. *Plant Journal*, **18**, 193–201.

Ausin, I., Alonso-Blanco, C., Jarillo, J.A., Ruiz-Garcia, L. and Martinez-Zapater, J.M. (2004). Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nature Genetics*, **36**(2), 162-166.

Bäurle, I., Smith, L., Baulcombe, D.C. and Dean, C. (2007). Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing *Science*, **318**(5847), 109-112.

Bäurle, I. and Dean, C. (2008). Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of Arabidopsis targets. *PLoS ONE*, **3**(7), e2733.

Bennett, M.D. and Leitch, I.J. (1995) Nuclear DNA amounts in angiosperms. *Annals of Botany*, **76**, 113-176.

Bennett, M.H., Mansfield, J.W., Lewis, M.J. and Beale, M.H. (2002). Cloning and expression of sesquiterpene synthase genes from lettuce (Lactuca sativa L.). *Phytochemistry*, **60**(3), 255-261.

Blackman, B.K., Strasburg, J.L., Raduski, A.R., Michaels, S.D. and Rieseberg, L.H. (2010). The Role of Recently Derived FT Paralogs in Sunflower Domestication. *Current Biology*, **20**(7), 629-635.

Blazquez, M.A., Ahn, J.H. and Weigel, D. (2003). A thermosensory pathway controlling flowering time in Arabidopsis thaliana. *Nature Genetics*, **33**(2), 168-171.

Bohn G. W. and Whitaker T.W. (1951) Recently introduced varieties of head lettuce and methods used in their development. *Agric. Circ.* 881.

Bodt, S.D., Raes, J., Peer, Y.V. and Theissen, G. (2003). And then there were many: MADS goes genomic. *Trends in Plant Science*, **8**, 475-483.

Borden, K.L.B. (1998). RING fingers and B-boxes: Zinc-binding protein-protein interaction domains. *Biochemistry and Cell Biology*, **76**(2-3), 351-358.

Boss, P.K., Bastow, R.M., Mylne, J.S. and Dean, C. (2004). Multiple pathways in the decision to flower: Enabling, promoting, and resetting. *Plant Cell*, **16**(SUPPL.).

Carmona, M.J., Cubas P., Calonje, M. and Martinez-Zapater, J.M. (2007). Flowering transition in grapevine (Vitis vinifera L.). *Canadian Journal of Botany*, **85**, 701–711.

Cashmore, A.R., Jarillo, J.A., Wu, Y-J. and Liu, D. (1999). Cryptochromes: Blue Light Receptors for Plants and Animals. *Science*, **284**(5415), 760-765.

Cerdan, P.D. and Chory, J. (2003). Regulation of flowering time by light quality. *Nature*, **423**(6942), 881-885.

Chia T.Y.P., Muller A., Jung C. and Mutasa-Gottgens E.S. (2008). Sugar beet contains a large *CONSTANS-LIKE* gene family including a *CO* homologue that is independent of the early-bolting (*B*) gene locus. *Journal of Experimental Botany*, **59**, 2735-2748.

Choi J., Hyun Y., Kang M.J., Yun H, Yun J.Y., Lister C., Dean C., Amasino R.M., Noh B., Noh Y.S. and Choi Y. (2009). Res**et**ting and regulation of FLOWERING LOCUS C expression during *Arabidopsis* reproductive development. *Plant Journal*, **57**, 918–931.

Clamp, M., Cuff, J., Searle, S.M. and Barton, G.J. (2004). The Jalview Java alignment editor. *Bioinformatics*, **20**, 426-427.

Clarke, J.H. and Dean, C. (1994). Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *Molecular and General Genetics*, **242**(1), 81-89.

Clough, S.J. and Bent, A.F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, **16**, 735-743.

Coles, J.P., Phillips, A.L., Croker, S.J., García-Lepe, R., Lewis, M.J. and Hedden, P. (1999). Modification of gibberellin production and plant development in *Arabidopsis* by sense and antisense expression of gibberellin 20-oxidase genes. *Plant Journal*, **17,** 547–556.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R. and Henikoff, S. (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant Journal*, **37**(5), 778-786.

Corbesier, L. and Coupland, G. (2005). Photoperiodic flowering of Arabidopsis: Integrating genetic and physiological approaches to characterization of the floral stimulus. *Plant, Cell and Environment*, **28**(1), 54-66.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C. and Coupland, G. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis, *Science*, **316(**5827), 1030-1033.

Cullings, K.W. (1992). Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology*, **1**, 233-240.

Curtis, I.S., Nam, H.G., Yun, J.Y. and Seo, K.H. (2002). Expression of an antisense GIGANTEA (GI) gene fragment in transgenic radish causes delayed bolting and flowering, *Transgenic Research*, **11**, 249-256.

Curtis, I.S., Power, J.B., Blackhall, N.W., De Laat, A.M.M. and Davey, M.R. (1994). Genotype-independent transformationof lettuce using Agrobacterium tumefaciens. *Journal of Experimental Botany*, **45**, 1441–144.

Dalmais, M., Schmidt, J., Le Signor, C., Moussy, F., Burstin, J., Savois, V., Aubert, G., Brunaud, V., de Oliveira, Y., Guichard, C., Thompson, R., and Bendahmane, A. (2008). UTILLdb, a *Pisum sativum in silico* forward and reverse gen**et**ics tool. *Genome Biology*, **9**, R43.

Decousset, L., Griffiths, S., Dunford, R.P., Pratchett, N. and Laurie D.A. (2000) Development of STS markers closely linked to the *Ppd-H1* photoperiod response gene of barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **101**:1202-1206.

Department for Environment, Food and Rural Affairs (**DEFRA) Basic Horticultural Statistics.** (2009) – [online]. Available at; http://statistics.defra.gov.uk/esg/.

De Vries, I.M. (1990). Crossing experiments of lettuce cultivars and species (Lactuca sect. Lactuca, Compositae). *Plant Systematics and Evolution*, **171**, 233-248.

De Vries, I.M. (1997). Origin and domestication of Lactuca sativa L. *Genetic Resources and Crop Evolution*, **44**(2), 165-164.

Devlin, P.F. and Kay, S.A. (2000). Flower arranging in Arabidopsis. *Science*, **288**(5471), 1600-1602.

Dielen, V., Notte, C., Lutts, S., Debavelaere, V., Van Herck, J.C. and Kinet, J.M. (2005). Bolting control by low temperatures in root chicory (Cichorium intybus var. sativum) *Field Crops Research*, 76-85.

Doležalová, I., Lebada, A., Janeček, J., Číhalíková, J., Křístková, E. and Vránová, O. (2002). Variation in chromosome numbers and nuclear DNA contents in genetic resources of *Lactuca* L. species (*Asteraceae*). *Genetic Resources and Crop Evolution.* **49**. 383-395.

Doyle, J.J. and and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, **19**, 11-15.

Dubnau, J. and Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. *Nature*, **379**, 694–699.

Dunford, R.P., Griffiths, S., Christodoulou, V. and Laurie, D.A. (2005). Characterisation of a barley (*Hordeum vulgare L.*) homologue of the *Arabidopsis* flowering time regulator *GIGANTEA*. *Theoretical and Applied Genetics*, **110**, 925-931.

Edwards, K.D., Lynn, J.R., Gyula, P., Nagy, F. and Millar, A.J. (2005). Natural Allelic Variation in the Temperature-Compensation Mechanisms of the Arabidopsis thaliana Circadian Clock. *Genetics*, **170**(1), 387-400.

El-Assal, S.E.D., Alonso-Blanco, C., Peeters, A.J.M., Wagemaker, C., Weller, J. L. and Koornneef, M. (2003). The role of cryptochrome 2 in flowering in Arabidopsis. *Plant Physiology*, **133**, 1504-1516.

Endo, M., Mochizuki, N., Suzuki, T. and Nagatani, A. (2007). CRYPTOCHROME2 in Vascular Bundles Regulates Flowering in *Arabidopsis*. *Plant Cell*, **19**, 84-93.

Faure, S., Higgins, J., Turner, A. and Laurie, D.A. (2007). The *FLOWERING LOCUS T*-like gene family in barley (*Hordeum vulgare*). *Genetics*, **176**, 599-609.

Food and Agriculture Organisation of the United Nations-Agricultural Data (**FAOSTAT).** (2009). [online]. Available at: http://faostat.fao.org/site/291/default.aspx.

Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N, and Banfield, M.J. (2003). *DETERMINATE* and *LATE FLOWERING* are two *TERMINAL FLOWER1/CENTRORADIALIS* homologs that control two distinct phases of flowering initiation and development in pea. *Plant Cell*, **15**, 2742–2754.

Fritjers, A.C.J., Zhang, Z., van Damme, M., Wang, G.-L., Ronald, P.C. and Michelmore, R.W. (1997). Construction of a bacterial artificial chromosome library containing large EcoRI and HindIII genomic fragments of lettuce. *Theoretical Applied Genetics*, **94**, 390-399.

Funk, V.A., Bayer, R.J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., Schilling, E., Panero, J.L., Baldwin, B.G., Garcia-Jagas, N., Susanna, A. and Jansen, R.K. (2005). Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter*, **55**: 343-374.

Gendall, A.R., Levy, Y.Y., Wilson, A. and Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*, **107**(4), 525-535.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C,. Burtner, C., Johnson, J.E. and Odden A.R. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics*, **164**, 731–740.

Griffiths, S., Dunford, R.P., Coupland, G. and Laurie, D.A. (2003). The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. *Plant Physiology*, **131**, 1855-1867.

Grishin, N. (2001). KH domain: one motif, two folds. *Nucleic Acids Research*, **29**, (3) 638-643.

Gu, Q., Ferrándiz, C., Yanofsky, M.F., Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development*, **125**, 1509-1517.

Guo, H., Yang, H., Mockler, T.C. and Lin, C. (1998). Regulation of Flowering Time by Arabidopsis Photoreceptors. *Science*, **279**(5355), 1360-1363.

Hakimi, M.A., Dong, Y., Lane, W.S., Speicher, D.W. and Shiekhattar, R. (2003). A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *The Journal of Biological Chemistry*, **278**, 7234–7239.

Halliday, K.J., Salter, M.G., Thingnaes, E. and Whitelam, G.C. (2003). Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator FT. *Plant Journal*, **33**(5), 875-885.

Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000). An RNA directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, **404**, 293-296.

Hanzawa, Y., Money, T. and Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. *Proceedings of the National Academy of Science*, **102**, 7748–7753.

Harlan, J.R. (1986) Lettuce and Sycamore: Sex and romance in Ancient Egypt. *Economic Botany* 40 (1) 4-15

Haughn, G. and Gilchrist, E. (2006). TILLING in the Botanical Garden: A Reverse Genetic Technique Feasible for all Plant Species. *Floriculture, Ornamental and Plant Biotechnology*, **1**, 476-482.

Hayama, R., Yokoi, S., Tamaki, S., Yano, M. and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature*, **422**, 719-722.

He, Y., Michaels, S.D. and Amasino, R.M. (2003). Regulation of Flowering Time by Histone Acetylation in Arabidopsis. *Science*, **302**(5651), 1751-1754.

Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J.,Vardy, M.E., Ellis, N., Beltran, J.P., Rameau, C. and Weller, J.L. (2005). Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiology*, **137**, 1420-1434.

Hecht, V., Knowles, C.L., Vander Schoor, J.K., Lim, C.L., Jones, S.E., Lambert, M.J.M. and Weller, J.L. (2007). Pea *Late Bloomer1* is a *Gigantea* ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant Physiology*, **144**, 648-661.

Higgins, J.A., Bailey, P.C. and Laurie, D.A. (2010). Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. *PLoS*, 5(4), e10065.

Hohmann, U., Jacobs, U. and Jung, G. (2005). An EMS mutagenesis protocol for sugar beet and isolation of non-bolting mutants. *Plant Breeding*, **124**, 317-321.

Hornyik, C., Terzi, L.C., and Simpson, G.G. (2010). The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Developmental Cell*, **18**, 203–213.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Einchholtz, D., Rogers, S.G. and Fraley, R.T. (1985). A simple and general method for transferring genes into plants. *Science*. **227**, 1229–1231.

Hruz, T., Caule, O., Szabo, G., Wessendorf, F., Bleulers, S., Oertie, L., Widmayer, P., Gruissem, W., Zimmermann, P. (2008). Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics*, Article ID: 420747.

Huang, S.S., Raman, A.S., Ream, J.E., Fujiwara, H., Cerny, R.E. and Brown, S.M. (1998). Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in *Arabidopsis*. *Plant Physiology*, **118**, 773–781.

Humphrey, G.W., Wang, Y., Russanova, V.R., Hirai, T., Qin, J., Nakatani, Y. and Howard, B.H. (2001). Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *Journal of Biological Chemistry*, **276**, 6817–6824.

Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A. and Kay, S.A. (2005). Plant science: FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science,* **309**(5732), 293-297.

Izawa, T., Oikawa, T., Sugiyama, N., Tanisaka, T., Yano, M., and Shimamoto, K. *(*2002). *Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. Genes & Development*, **16**, 2006*–2020.*

Izawa, T., Takahashi, Y. and Yano, M. (2003). Comparative biology comes into bloom: Genomic and genetic comparison of flowering pathways in rice and Arabidopsis. *Current Opinion in Plant Biology*, **6**(2), 113-120.

Jackson, S.D. (2009). Plant responses to Photoperiod. *New Phytologist*, **181**, 517-531.

Jaeger, K.E. and Wigge, P.A. (2007). FT Protein Acts as a Long-Range Signal in *Arabidopsis*. *Current Biology*, **17**, 1050-1054.

Jang, S., Marchal, V., Panigrahi, K.C., Wenkel, S., Soppe, W., Deng, X.W., Valverde, F. and Coupland, G. (2008). *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *EMBO Journal*, **27**, 1277-1288.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science,* **290**(5490), 344-347.

Jung, J.-H., Seo, Y.-H., Seo, P.J., Reyes, J.L., Yun, J., Chua, N.-H. and Park, C.-M. (2007). The *GIGANTEA*-Regulated MicroRNA172 Mediates Photoperiodic Flowering Independent of *CONSTANS* in *Arabidopsis*. *Plant Cell*, **19**, 2736-2748.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J. and Weigel, D. (1999). Activation Tagging of the Floral Inducer FT. *Science,* **286**(5446), 1962-1965.

Kesseli, R.V., Ochoa, O. and Michelmore, R.W. (1991) Variation at RFLP loci in *Lactuca* spp. and the origin of cultivated lettuce (*L. sativa*). *Genome,* **34,** 430-436.

Kesseli, R.V., Paran, I. and Michelmore, R.W. (1994). Analysis of a detailed genetic linkage map of Lactuca sativa (lettuce) constructed from RFLP and RAPD markers. *Genetics*, **136**(4), 1435-1446.

Kim, H.J., Hyun, Y., Park, J.Y., Park, M.J., Park, M.K., Kim, M.D., Kim, H.J., Lee, M.H., Moon, J., Lee, I. and Kim, J. (2004). A genetic link between cold responses and flowering time through FVE in Arabidopsis thaliana. *Nature Genetics,* **36**(2), 167-171.

Kim, J., Kim, Y., Yeom, M., Kim, J.-H. and Nam, H.G. (2008). *FIONA1* Is Essential for Regulating Period Length in the *Arabidopsis* Circadian Clock. *Plant Cell*, **20**, 307-319.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T. and Yano, M. (2002). Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant and Cell Physiology*, **43**(10), 1096-1105.

Koopman, W.J.M., De Jong, J.H. and De Vries, I.M. (1993) Chromosome banding in lettuce species (*Lactuca* sect. *Lactuca*, *Compositae*). *Pl. Syst. Evol.*, 185, 249-257.

Koopman, W.J.M. and De Jong, J.H. (1996). A numerical analysis of karyotypes and DNA amounts in lettuce cultivars and species (Lactuca subsect. Lactuca, Compositae). *Acta Botanica Neerlandica*, **45**(2), 211-222.

Koornneef, M., Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart C.J. and Peeters, A.J.M. (1998). Genetic Interactions Among Late-Flowering Mutants of Arabidopsis**.** *Genetics*, **148**(2), 885–892.

Koornneef, M., Hanhart, C.J. and Van Der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. *Molecular and General Genetics*, **229**(1), 57-66.

Krieg, D.R. (1963). Ethyl methanesulfonate-induced reversion of bacteriophage T4*r*II mutants. *Genetics*, **48,** 561–580.

Langridge, J. (1957). Effect of Day-Length and Gibberellic Acid on the Flowering of Arabidopsis. *Nature*, **180**, 36-37.

Ledger, S., Strayer, C., Ashton, F., Kay, S.A. and Putterill, J. (2001). Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. *Plant Journal*, **26**, 15.

Lee, I., Aukerman, M.J., Gore, S.L., Lohman, K.N., Michaels, S.D., Weaver, L.M., John, M.C., Feldmann, K.A. and Amasino, R.M. (1994). Isolation of LUMINIDEPENDENS: A Gene Involved in the Control of Flowering Time in Arabidopsis. *The Plant Cell*, **6**(1), 75-83.

Lee, J.H., Cho, Y.S., Yoon, H.S., Suh, M.C., Moon, J. and Lee, I. (**2005).** Conservation and divergence of FCA function between Arabidopsis and rice. *Plant Molecular Biology*, **58**, 823–838.

Levy, Y.Y. and Dean, C. (1998). The transition to flowering. *Plant Cell*, **10**(12), 1973-1989.

Levy, Y.Y., Mesnage, S., Mylne J.S., Gendall, A.R. and Dean, C. (2002) Multiple roles of *Arabidopsis* in vernalisation and flowering time control. *Science*, **297**, 243-6.

Li, D., and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Sciences*, **58**, 2085–2097

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078-9.

Lim, M.-H., Kim, J., Kim, Y.-S., Chung, K.-S., Seo, Y.-H., Lee, I., Kim, J., Hong, C. B., Kim, H.-J. and Park, C.-M. (2004). A New Arabidopsis Gene, FLK, Encodes an RNA Binding Protein with K Homology Motifs and Regulates Flowering Time via FLOWERING LOCUS C. *The Plant Cell*, **16**(3), 731-740.

Lindqvist, K. (1960a). On the origin of cultivated lettuce. *Hereditas*, **46**, 319-350.

Lindqvist, K. (1960b). Cytogenetic studies of the *serriola* group of *Lactuca*. Hereditas **46**. 75-151

Liu, C., Chen, H., Er, H.L., Soo, H.M. and Kumar, P.P. (2008). Direct interaction of AGL24 and SOC1 integrates flowering signals in *Arabidopsis*. *Development*, **135**, 1481–91.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S. and Dean C. (2010). Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science*, **327**, 94-97.

Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S. and Dean, C. (2007). The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Molecular Cell*, **28**, 398–407.

Logemann, J., Schell, J. and Willmitzer, L. (1987). Improved method for the isolation of RNA from plant tissues. *Analytical Biochemistry*, **163**(1), 16-20.

Lu, R., Martin-Hernandez, A.M., Peart, J.R., Malcuit, I. and Baulcombe, D.C. (2003). Virus-induced gene silencing in plants. *Methods*, **30**, 296-303.

Lucas, H., Feuerbach, F., Kunert, K., Grandbastien, M.A. and Caboche, **M.** (1995). RNA-mediated transposition of the tobacco retrotransposon Tnt1 in Arabidopsis thaliana. *EMBO Journal*, **14**(10), 2364-2372.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. and Dean, C. (1997). FCA, a Gene Controlling Flowering Time in Arabidopsis, Encodes a Protein Containing RNA-Binding Domains. *Cell*, **89**(5), 737-745.

Macknight, R., Duroux, M., Laurie, R., Dijkwel, P., Simpson, G. and Dean, C. (2002). Functional significance of the alternative transcript processing of the Arabidopsis floral promoter *FCA*. *Plant Cell*, **14,** 877–888.

Makeyev, A. and Liebhaber, S.A. (2002). The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *RNA*, **8,** 265–278

Manzano, D., Marquardt, S., Jones, A.M.E., Bäurle, I., Liu, F. and Dean, C. (2009) Altered interactions within FY/AtCPSF complexes required fro Arabidopsis FCA-mediated chromatin silencing. PNAS, **106**(21), 8772-8777

Marquardt, S., Boss, P.K., Hadfield, J. and Dean, C. (2006). Additional targets of the Arabidopsis autonomous pathway members, FCA and FY. *Journal of Experimental Botany*, **57**(13), 3379-3386.

Mas, P., Kim, W.-Y., Somers, D.E. and Kay, S.A. (2003). Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. *Nature*, **426**, 567-570.

Mathieu, J., Warthmann, N., Küttner, F. and Schmid, M. (2007). Export of FT protein from phloem companion cells Is sufficient for floral induction in *Arabidopsis*. *Current Biology*, **17**, 1055-1060.

Mazier, M., Botton, E., Flamain, F., Bouchet, J.P., Courtial, B., Chupeau, M.C., Chupeau, Y., Maisonneuve, B. and Lucas, H. (2007). Successful gene tagging in lettuce using the Tnt1 retrotransposon from tobacco. *Plant Physiology*, **144**(1), 18-31.

McClung, C. (2006). Plant Circadian Rhythms. *The Plant Cell*, **18**(4), 792-803.

McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000), Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Nature Biotechnology*, **18**, 455-457.

McGuire, P.E., Ryder, E.J., Clark, R.W., Antle, R., Emery, G. and Hannan, R.M. (1993). Genetic Resources of Lettuce and Lactuca species in California. An assessment of the USDA and UC Collections and Recommendations for Long-term Security. University of California Genetic Resources Conservation Program, Davis, USA. Report No. 12.

Mead, R., Curnow, R.N. and Hasted, A.M. (1993). Statistical Methods in Agriculture and Experimental Biology, Second edition. London: Chapman & Hall.

Michaels, S.D. and Amasino, R.M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**(5), 949-956.

Michaels, S.D. and Amasino, R.M. (2000). Memories of winter: vernalization and the competence to flower. *Plant Cell and Environment*, **23**, 1145-1154.

Michaels, S.D. and Amasino, R.M. (2001). Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGADA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell*, **13**(4), 935-941.

Michaelson, M.J., Price, H.J., Ellison, J.R. and Johnston, J.S. (1991). Comparison of plant DNA contents determined by feulgen microspectrophotometry and laser flow cytometry. *American Journal of Botany*, **78**, 183-188.

Michelmore, R. (2007). White paper: Priorities for research, education and extension in genomics, genetics, and breeding of the Compositae.

Minoia, S., Petrozza, A., D'Onofrio, O., Piron F., Mosca G., Sozio G., Cellini F., Bendahmane A. and Carriero F. (2010). A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Research Notes*, **3**, 69.

Mitchell, P.J. and Tijan, R. (1989). Transcriptional regulation in mammalian cells by sequence specific DNA binding proteins. *Science*, **245**, 371-378.

Mockler, T.C., **Yu**, **X. and Shalitin, D.** (2004). Regulation of flowering time in *Arabidopsis* by K homology domain proteins. *Proceedings of the National Academy of Sciences*, **101**, 12759–12764.

Morel, P., Tréhin, C., Breuil-Broyer, S. and Negrutiu, I. (2009). Altering FVE/MSI4 results in a substantial increase of biomass in Arabidopsis - the functional analysis of an ontogenesis accelerator. *Molecular Breeding*, **23,** 239–257.

Mouradov, A., Cremer, F. and Coupland, G. (2002). Control of flowering time: Interacting pathways as a basis for diversity. *Plant Cell*, **14,** (SUPPL.).

Nakasako, M., Matsuoka, D., Zikihara, K. and Tokutomi, S. (2005). Quaternary structure of LOV-domain containing polypeptide of Arabidopsis FKF1 protein. *FEBS Letters*, **579**(5), 1067-1071.

Napp-Zinn, K. (1987). Vernalization: Environmental and genetic regulation. *In Manipulation of Flowering*. Edited by Atherton, J.G. Butterworths, London, 123-132

NCBI (National Center for Biotechnology Information). (2008). BLAST: BasicLocal Alignment Search Tool. [online]. Available at: http://blast.ncbi.nlm.nih.gov/Blast.cgi.

Nelson, D.C., Lasswell, J., Rogg, L.E., Cohen, M.A. and Bartel, B. (2000). FKF1, a clock-controlled gene that regulates the transition to flowering in Arabidopsis. *Cell*, **101**(3), 331-340.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). In *Applied linear statistical models* 4th Edition. McGraw-Hill.

Nishioka, M., Tamura, K., Hayashi, M., Fujimori, Y., Ohkawa, Y., Kuginuki, Y. and Harada, K. (2005). Mapping of QTLs for bolting time in Brassica rapa (syn. campestris) under different environmental conditions. *Breeding Science*, **55**(2), 127-133.

Oliver, G.W. (1910). New Methods of Plant Breeding. *Bureau of Plant Industry Bulletin, U.S. Department of Agriculture*, **167**.

Oyama, T., Shimura, Y. and Okada, K. (1997). The *Arabidopsis HY5* gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes and Development*, **11**, 2983-2995.

Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A. and Nam, H.G. (1999). Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis GIGANTEA* gene. *Science*, **285**, 1579-1582.

Patterson, H.D., Williams, E.R. and Hunter, E.A. (1978). Block designs for variety trials. *Journal of Agricultural Science*, **90**, 395-400.

Perrière, G. and Gouy, M. (1996). WWW-Query: An on-line retrieval system for biological sequence banks. *Biochimie*, **78**, 364-369.

Porebski, S., Bailey, L.G. and Baum, B.R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, **15**, 8-15.

Pink, D.A.C. and Keane, E.M. (1993). Lettuce *Lactuca sativa* L. In *Genetic Improvement of Vegetable Crops*. Edited by B. Kalloo and B.O. Bergh Oxford, Pergamon Press**,** 543-571.

Prince, S.D., Marks, M.K. and Carter, R.N. (1979). Induction of Flowering in Wild Lettuce (Lactuca serriola L.). *New Phytologist*, **81**, 265-277.

Putterill, J., Robson, F., Lee, K., Simon, R. and Coupland, G. (1995). The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell*, **80**(6), 847-857.

Quail, P.H. (2002). Photosensory perception and signalling in plant cells: New paradigms? *Current Opinion in Cell Biology*, **14**(2),180-188.

Ratcliffe, O. J., Kumimoto, R. W., Wong, B. J. and Riechmann, J. L. (2003). Analysis of the Arabidopsis MADS AFFECTING FLOWERING Gene Family: MAF2 Prevents Vernalization by Short Periods of Cold. *The Plant Cell*, **15**(5), 1159-1169.

Redei, G.P. (1962). Supervital mutants of Arabidopsis. *Genetics*, **47**, 443-460.

Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Pineiro, M., Reeves, P.H., Putterill, J. and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant Journal*, **28**(6), 619-631.

Robinson, R.W. (1986). Mutagenesis of lettuce with ethyl methane sulfonate. *Mutation Breeding Newsletter*, **28**, 7.

Rodenburg, C.M. (1960). Varieties of lettuce; an international monograph. Tjeenk Willink, Zwolle: 260.

Rozen, S. and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386

Rulkens A.J.H. (1987). CGN report: CGN-T Wageningen 51.

Ryder, E.J. (1983). Inheritance, linkage, and gene interaction studies in lettuce. *Journal of the American Society for Horticultural Science*, **108**, 985-991.

Ryder, E.J. (1985). Use of early flowering genes to reduce generation time in backcrossing, with specific application to lettuce breeding. *Journal of the American Society for Horticultural Science*, **110**, 570-573.

Ryder, E.J. (1986). Lettuce breeding In *Breeding Vegetable Crops*, Edited by Bassett M. AVI Publishing, Westport, Connecticut, USA.

Ryder, E.J. (1999). Evolution and Crop History. *Lettuce, Endive and Chicory*. Wallingford, CABI Publishing**:** 15-23.

Ryder, E.J. and Johnson, A.S. (1974). Mist depollination of lettuce flowers. *HortScience,* **9**, 584.

Ryder, E.J. and Milligan, D.C. (2005). Additional genes controlling flowering time in *Lactuca sativa* and *L. serriola*. *Journal of the American Society for Horticultural Science*, **130**(3), 448-453.

Ryder E.J. and Waycott W. (1994). Crisphead lettuce resistant to corky root: cultivars 'Glacier' and 'Misty Day' and 16 resistant breeding lines. *HortScience,* **29**, 335-336.

Ryder E.J. and Whitaker T. (1976) Lettuce, Lactuca sativa (Compositae). In N.W. Simmons (Ed.). Evolution of crop plants. Longman, London, 39-41.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. and Coupland, G. (2000). Distinct roles of constans target genes in reproductive development of Arabidopsis. *Science*, **288**(5471), 1613-1616.

Sanda, S.L. and Amasino, R.M. (1996). Interaction of *FLC* and late-flowering mutations in *Arabidopsis thaliana*. *Molecular and General Genetics,* **251**, 69-74.

Sawa, M., Nusinow, D.A., Kay, S.A. and Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science*, **318**, 261-265.

Schomburg, F.M., Patton, D.A., Meinke, D.W. and Amasino, R.M. (2001). FPA, a Gene Involved in Floral Induction in Arabidopsis, Encodes a Protein Containing RNA-Recognition Motifs. *The Plant Cell*, **13**(6), 1427-1436.

Scortecci, K.C., Michaels, S.D. and Amasino, R.M. (2001). Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. *The Plant Journal*, **26**, 229–236.

Searle, I. and Coupland, G. (2004). Induction of flowering by seasonal changes in photoperiod. *EMBO Journal*, **23**(6), 1217-1222.

Sessa, R.A., Bennett, M.H., Lewis, M.J., Mansfield, J.W. and Beale, M.H. (2000). Metabolite profiling of sesquiterpene lactones from lactuca species: Major latex components are novel oxalate and sulfate conjugates of lactucin and its derivatives. *Journal of Biological Chemistry*, **275**(35), 26877-26884.

Shannon, S. and Meeks-Wagner, D.R. (1991). A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development. *Plant Cell,* **3**(9), 877-892.

Shindo C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordberg, M. and Dean C. (2005) Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of Arabidopsis. Plant Physiology, **138**, 1163-1173.

Shindo C., Lister, C., Crevillen, P., Nordberg, M. and Dean, C. (2006). Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalisation response. *Genes and Development*, **20**, 3079-3083.

Simpson, G.G. (2004). The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Current Opinion in Plant Biology,* **7**, 570–74

Simpson, G.G. and Dean, C. (2002). Arabidopsis, the Rosetta stone of flowering time? *Science,* **296**(5566), 285-289.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I. and Dean, C. (2003). FY Is an RNA 3' End-Processing Factor that Interacts with FCA to Control the Arabidopsis Floral Transition. *Cell*, **113**(6), 777-787.

Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. and Facciotti, D. (2005). A reverse genetic, non-transgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology*, **23**, 75-81.

Soundy, P. and Smith, I.E. (1992). Response of lettuce (*Lactuca sativa* L) to nitrogen and phosphorus fertilization. *Journal of the Southern African Society for Horticultural Sciences*, **2**, 82–85.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, **98**(3), 503-517.

Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps, J.A. and Kay, S.A. (2000). Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science*, **289**(5480), 768-771.

Streitner, C., Danisman, S., Wehrle, F., Schöning, J.C., Alfano, J.R. and Staiger, D. (2008). The small glycine-rich RNA binding protein *At*GRP7 promotes floral transition in *Arabidopsis thaliana*. *The Plant Journal*, **56**(2), 239–250.

Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F. and Coupland, G. (2001). *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*, **410**, 1116-1120.

Sung, S. and Amasino R.M. (2004) Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature*, **427**, 159–164.

Sung, S., Schmitz, R.J. and Amasino, R.M. (2006). A PHD finger protein involvedin both the vernalization and photoperiod pathways in Arabidopsis. *Genes and Development*, **20**, 3244–3248.

Syed, N.H. and Flavell, A.J. (2007). Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. *Nature Protocols*, **1**(6), 2746-2752.

TAIR (The Arabidopsis Information Resource). (2009). Arabidopsis germplasm collection. Available at: http://www.arabidopsis.org.

Tamaki, S., Matsuo, S., Hann, L.W., Yokoi, S. and Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science*, **316**(5827), 1033-1036.

Thomas, B., Carre, I. and Jackson, S. (2006). Photoperiodism and Flowering. In *The Molecular Biology and Biotechnology of Flowering.* Edited by B. R. Jordan. Wallingford, CABI Publishing.

Thompson, R.C., Doolittle, S.P. and Henneberry. T.J. (1958). Growing lettuce in greenhouses. U.S. Dept. of Agriculture in Washington, D.C.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876-4882.

Till, B.J., Burtner, C., Comai, L. and Henikoff, S. (2004). Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research*, **32**(8), 2632-2641.

Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S. and Comai, L. (2007). **Discovery of chemically induced mutations in rice by TILLING.** *BMC Plant Biology*, **7,** 19.

Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L. and Henikoff, S. (2003). Large-Scale Discovery of Induced Point Mutations With High-Throughput TILLING. *Genome Research*, **13**(3), 524-530.

Timms, L., Jimenez, R., Chase, M., Lavelle, D., McHale, L., Kozik, A., Lai, Z., Heesacker, A., Knapp, S., Rieseberg, L., Michelmore, R. and Kesseli, R. (2006). Analyses of synteny between Arabidopsis thaliana and species in the asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. *Genetics*, **173**(4), 2227-2235.

Torii K.U., Stoop-Myer, C.D., Okamoto, H., Coleman, J.E., Matsui, M., and Deng, X.W. (1999). *The RING finger motif of photomorphogenic repressor COP1 specifically interacts with the RING-H2 motif of a novel Arabidopsis protein. Journal of Biological Chemistry*, **274**, 27674-*27681*.

Trick, M., Long, Y., Meng, J. and Bancroft, I. (2009). Single nucleotide polymorphism (SNP) discovery inthe polyploid *Brassica napus* using Solexa transcriptome sequencing**.** *Plant Biotechnology Journal*, **7**, 334–346.

Truco, M.J., Antonise. R., Lavelle, D., Ochoa, O., Kozik, A., Witsenboer, H., Fort, S.B., Jeuken, M.J., Kesseli, R.V., Lindhout, P., Michelmore, R.W. and Peleman J. (**2007**). A high-density, integrated genetic linkage map of **lettuce** (*Lactuca* spp.). *Theoretical and Applied Genetics*, **115**, 735–746.

Turner, A., Beales, J., Faure, S., Dunford. and Laurie, D. (2005). The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in Barley. *Science*, **310**, 1031-1034.

Valverde, R., Edwards, L. and Regan, L. (2008). Structure and function of KH domains. *FEBS Journal*, **275**(11), 2712–2726,

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A. and Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science*, **303**, 1003-1006.

van Nockler, S., Muszynski, M., Briggs, K. and Amasino, R. (2000). Characterization of a gene from Zea mays related to the Arabidopsis flowering-time gene LUMINIDEPENDENS. *Plant Molecular Biology*, **44**, 107-122.

Veley, K.M. and Michaels, S.D. (2008). Functional redundancy and new roles for genes of the autonomous floral-promotion pathway. *Plant Physiology*. **147**, 682–695.

Wilk, M.B. & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, **55**, 1-17.

Wilson, R.N., Heckman, J.W. and Somerville, C.R. (1992). Gibberellin is required for flowering in Arabidopsis thaliana under short days. *Plant Physiology*, **100**(1), 403-408.

Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., and Helliwell, C.A. (2006) The Arabidopsis thaliana vernalisation response requires a polycomb-like protein complex that also includes VENALISATION INSENSITIVE 3. Proc. Natl. Acad. Sci., **103**(39), 14631-6

Wroblewski, T., Piskurewicz, U., Tomczak, A., Ochoa, O. and Michelmore, R.W. (2007). Silencing of the major family of NBS-LRR-encoding genes in lettuce results in the loss of multiple resistance specificities. *The Plant Journal*, **51**, 803–818.

Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M. and Araki, T. (2005). *TWIN SISTER of FT* (*TSF*) acts as a floral pathway integrator redundantly with *FT*. *Plant and Cell Physiology*, **46**, 1175-1189.

Yang, C-H. and Chou, M-L. (1999). *FLD* Interacts with *CO* to Affect both Flowering Time and Floral Initiation in *Arabidopsis thaliana*. *Plant and Cell Physiology*, **40**,(6) 647-650.

Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S. and Ahn, J.H. (2005). *CONSTANS* activates *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in Arabidopsis. *Plant Physiology*, **139**, 770-778.

Yoo, S.K. Lee, J.S., and Ahn, J.H. (2006). Overexpression of AGAMOUS-LIKE 28 (AGL28) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. *Biochemical and Biophysical Research Communications*, **348,** 929–936.

Yoo, S.Y., Kardailsky, I., Lee, J.S., Weigel, D. and Ahn, J.H. (2004). Acceleration of flowering by overexpression of MFT (MOTHER OF FT AND TFL1). *Molecules And Cells*, **17**, 95-101.

APPENDIX I

Table A1 - Details of lines making up the wild lettuce diversity set

Origin of seed	Accession name/number	Species	Origin of seed	Accession name/number	Species	Origin of seed	Accession name/number	Species
USDA	Dehesa (W33)	*L. serriola*	**USDA**	PI-261653 (W8)	*L. saligna*	**USDA**	01G-158-VIR (W7)	*L. virosa*
USDA	Montalban (W34)	*L. serriola*	**USDA**	HH-2350-1 (W36)	*L. saligna*	**USDA**	USDA-SAL-167 (W11)	*L. virosa*
USDA	Davis Wheat (W35)	*L. serriola*	**USDA**	HH-2351-1 (W37)	*L. saligna*	**USDA**	USDA-IVT-1398-SAL-177 (W12)	*L. virosa*
USDA	HH-2383 (W40)	*L. serriola*	**USDA**	HH-2352-4 (W38)	*L. saligna*	**USDA**	USDA_SAL-193 (W13)	*L. virosa*
USDA	HH-1408 (W41)	*L. serriola*	**USDA**	HH-2353-1 (W39)	*L. saligna*	**USDA**	USDA-IVT-280 (W15)	*L. virosa*
USDA	HH-2402 (W42)	*L. serriola*	**USDA**	00G-950 (W63)	*L. saligna*	**USDA**	USDA-SAL-175 (W18)	*L. virosa*
USDA	HH-2464 (W43)	*L. serriola*	**USDA**	CGN-5318 (W64)	*L. saligna*	**USDA**	USDA-SAL-180 (W20)	*L. virosa*
USDA	LSE57-15 (W44)	*L. serriola*	**USDA**	CGN-9311 (W65)	*L. saligna*	**USDA**	USDA-SAL-209 (W23)	*L. virosa*
USDA	UC96US23 (W45)	*L. serriola*	**USDA**	CGN-13330 (W66)	*L. saligna*	**USDA**	USDA-PI-273597 (W27)	*L. virosa*
USDA	USDA-SAL-28 (W46)	*L. serriola*	**USDA**	PI-491000 (W67)	*L. saligna*	**USDA**	USDA-PI-279138 (W30)	*L. virosa*
USDA	USDA-SAL-26 (W47)	*L. serriola*	**USDA**	PI-491226 (W68)	*L. saligna*	**USDA**	94G-173-9-green-virosa (W31)	*L. virosa*
USDA	UC99TRK-18-4 (W51)	*L. serriola*	**USDA**	PI-509523 (W69)	*L. saligna*	**USDA**	90G-248-red-virosa (W32)	*L. virosa*
USDA	02G993-2-PAVIA (W52)	*L. serriola*	**USDA**	PI-509525 (W70)	*L. saligna*	**HRI**	HRI1591	*L. virosa*
USDA	CGN-5916 (W54)	*L. serriola*	**USDA**	UC-1 (W71)	*L. saligna*	**HRI**	HRI1594	*L. virosa*
USDA	CGN-14263 (W55)	*L. serriola*	**USDA**	CGN-5157 (W72)	*L. saligna*	**HRI**	HRI1583	*L. virosa*
USDA	CGN-14278 (W57)	*L. serriola*	**USDA**	CGN-5267 (W73)	*L. saligna*	**HRI**	HRI1586	*L. virosa*
USDA	W66331A (W8)	*L. serriola*	**USDA**	CGN-5282 (W75)	*L. saligna*	**HRI**	HRI1589	*L. virosa*
USDA	LSE-102 (W60)	*L. serriola*	**USDA**	CGN-5309 (W76)	*L. saligna*	**HRI**	HRI1595	*L. virosa*
USDA	PIVT-1309 (W61)	*L. serriola*	**USDA**	CGN-5796 (W78)	*L. saligna*	**HRI**	HRI1597	*L. virosa*
USDA	LSE-18 (W62)	*L. serriola*	**USDA**	CGN-5882 (W79)	*L. saligna*	**HRI**	HRI1598	*L. virosa*

HRI	HRI1561	*L. serriola*	**USDA**	PI-491204_saligna (W80)	*L. saligna*	**HRI**	HRI1602	*L. virosa*
HRI	HRI1562	*L. serriola*	**USDA**	CGN-10888 (W81)	*L. saligna*	**HRI**	HRI6188	*L. virosa*
HRI	HRI1563	*L. serriola*	**USDA**	CGN-5147 (W83)	*L. saligna*	**HRI**	HRI6658	*L. virosa*
HRI	HRI1564	*L. serriola*	**USDA**	saligna_east_yolo (W84)	*L. saligna*	**HRI**	HRI6659	*L. virosa*
HRI	HRI1565	*L. serriola*	**USDA**	LJ-85314 (W85)	*L. saligna*	**HRI**	HRI6660	*L. virosa*
HRI	HRI1567	*L. serriola*	**USDA**	PI-491208 (W86)	*L. saligna*	**HRI**	HRI6661	*L. virosa*
HRI	HRI1569	*L. serriola*	**USDA**	CGN4662 (W87)	*L. saligna*	**HRI**	HRI6662	*L. virosa*
HRI	HRI1571	*L. serriola*	**USDA**	CGN9313 (W88)	*L. saligna*	**CGN**	CGN16198	*L. virosa*
HRI	HRI1572	*L. serriola*	**USDA**	CGN10883 (W89)	*L. saligna*	**CGN**	CGN16199	*L. virosa*
HRI	HRI1573	*L. serriola*	**USDA**	CGN5320 (W90)	*L. saligna*	**CGN**	CGN16200	*L. virosa*
HRI	HRI1575	*L. serriola*	**USDA**	CGN5329 (W91)	*L. saligna*	**CGN**	CGN16201	*L. virosa*
HRI	HRI1576	*L. serriola*	**USDA**	CGN5325 (W92)	*L. saligna*			
HRI	HRI1896	*L. serriola*	**USDA**	CGN5327 (W93)	*L. saligna*			
HRI	HRI1897	*L. serriola*	**USDA**	saligna_west_yolo (W94)	*L. saligna*			
HRI	HRI4978	*L. serriola*	**HRI**	HRI1625	*L. saligna*			
HRI	HRI5093	*L. serriola*	**HRI**	HRI1628	*L. saligna*			
HRI	HRI5094	*L. serriola*	**HRI**	HRI1629	*L. saligna*			
HRI	HRI5095	*L. serriola*	**HRI**	HRI1630	*L. saligna*			
HRI	HRI5499	*L. serriola*	**HRI**	HRI6382	*L. saligna*			
HRI	HRI5500	*L. serriola*	**HRI**	HRI6383	*L. saligna*			
HRI	HRI6187	*L. serriola*	**HRI**	HRI6384	*L. saligna*			
HRI	HRI1606	*L. serriola*	**HRI**	HRI6386	*L. saligna*			
HRI	HRI1624	*L. serriola*						
HRI	HRI6355	*L. serriola*						
HRI	HRI7141	*L. serriola*						

HRI	HRI7145	*L. serriola*
HRI	HRI13043	*L. serriola*

APPENDIX II

Primer sequences

All primers written 5'-3'

Chapter 2

EST sequencing and general sequencing primers

M13_F	CAGGAAACAGCTATGAC
M13_R	GTAAAACGACGGCCAG

Chapter 3

Isolation of a LsFT gene

FT_EST1_F1	GACCCGTTGGTTGTTGGAC
FT_EST1_R1	CTAGGACTTGGAGCATCAGGA
FT_exon4_F	TAGGGAGAGGGACCCGTTG
FT_exon4_R	TTATCTTCTTCGCCCACCAA

LsFT complementation studies

FT_comp_F	TCGTATGATGCCTAGGGAGAG
FT_comp_R	TTATCTTCTTCGCCCACCAA
Gateway_FT_ F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGTATG ATGCCTAGGGAGAG
Gateway_FT_ R	GGGGACACATTTGTACAAGAAAGCTGGGTTTATCTT CTTCGCCCACCAA
AtFT_comp_F	TTTGTTCAAGATCAAAGATGTCTA
AthFT_comp_R	AAGCCATCTAAAGTCTTCTTCCTC
AthFT_comp_att_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTTGTT CAAGATCAAAGATGTCTA
AthFT_comp_att_R	GGGGACCACTTTGTACAAGAAAGCTGGGTAAGCCA TCTAAAGTCTTCTTCC
FT_exon2-3_R	GGTCGCTGGTATATCGGTCA

Isolation of a LsFLK gene

FLK_EST1_F	TGTTGATTTCTGCAAAAGAGGA
FLK_EST1_R	TTTAATGGTGGCACCCTGTT
FLK_GW_GSP1	CGTCTCCTCAAGGTGGTTTCCCAATGA
FLK_GW_GSP2	GGGTGGTTTTGGGGCCAATCAGTTTAT
FLK_GW_seq_exon5_F	GCAATTCAAGAAACACGTGGA
FLK_intron5_F	TCCTGCTGCCGAGTATGG
FLK_R_intron2seqR	TGTCTTCTAATTGGAATCGAAATG
FLK_R_intron2_2seq	AAAAGATTGAAATGGATAAATCAAG

LsFLK complementation studies

FLK_comp_F	ATGGCAGAGGATAACATCAGTG
FLK_comp_R	TCAATACCCATAATTCCCTCCA
FLK_comp_att_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTA TGGCAGAGGATAACATCAGTG
FLK_comp_att_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTC AATACCCATAATTCCCTCCA
FLK_EXON2_F	CTGCCATTGATGGTTTGTTG
FLK_EST1_R	TTTAATGGTGGCACCCTGTT

Isolation of a LsFLD gene

FLD_EST_seq	TTCGCGATAAATGTCCACTC
FLD_EST_seq2	CGGTGAAACGTATCCTCCAA
FLD_GW_GSP1_F	TGCAAGACAACTGATGACATTTGGGTA
FLD_GW_GSP2_F	TAAGTTGCAGCTGCTGAT TTAGGAGGA
FLD_EST1_F	GTTGGTGGAATCGAACAGGT
FLD_EST1_R	CAAAATTGATGTATCCATGTGAGAG

FLD_5_GSP1	GAGAGTAAACCCGCGTTTGCATATTCT
FLD_5_GSP2	CTGCCCCTAGAGAAACATCTTGTGAGA
FLD_5_GSP1_2	TCGAGGGGT TCATGGTTGGAAATTTGA
FLD_5_GSP2_2	AGGCGCCAT GGACGATGATGATACATT
FLD_3_GSP1	TTCTCTTGGGTCGTATTCAAATGTTGC
FLD_3_GSP2	CGTTTTTAAGCGGGTTAAGAGAAGCTG
LsFLD complementation studies	
FLD_comp_F	ATGAACCCCTCGAACGAAAC
FLD_comp_R	AAAAATGTTAGTCCTTGAATATCACC
FLD_compATT_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAA CCCCTCGAACGAAAC
FLD_compATT_R	GGGGACCACTTTGTACAAGAAAGCTGGGTAAA AATGTTAGTCCTTGAATATCACC
FLD_exon1_F	GGCTCTAGTGGCAGGTGAAG
FLD_exon2_R	GAATACGACCCAAGAGAAAACG
Isolation of a LsLD gene	
LD_ESTseq	CAAGCTGCAATCGAAGAACA
LD_5UTR	CCCTCTACAAAAACAAAACCTCT
LD_5_R_seq2	GGGTTAATGATATGACGCTTCC
LD_RegI-II_F	TGGTGCTCTTTCTATTAAAATTGG
LD_BACseq_exon7	GAAGCCAAGCCATGAGAAAG
LD_GSP1_JAN10	CTTCAAGGTGGCGGGATAGGAGGAGAT
LD_GSP2_JAN10	TCACGTCACGCCCAAAACACATGAAAT
LD_3_GW_GSP1	GCGGAGGAACCTGTGTGGAAGAAGTGT
LD_3_GW_GSP2	GCTGTGTGCAGGTTCAATCGCAAATTC
LD_GSP1_1109	CCTGATCAAGCGAATCATTACGGGACA
LD_GSP2_1109	TGTCCATGTGACTGTAACCCGTTGGAGA
LD_GW_DEC09_GSP2	CCCAACAGCCAACTTCTTCTCACCAGA
LD_GW_DEC09_GSP1	TCAAAGATTCACCGATGTGGTCCCTGA
Isolation of a LsFVE gene	
FVE_ESTseqv2	TTGTTTGATCGTCGAAACCTT
FVE_exon7-8_F	AAATGCAGAATTTGCTCTTGC
FVE_exon7-8_R	TTGAATACTCCACAAAACAACTGA
FVE_5_GW_GSP1	GTTCAGAGAGGTAGAGTCTTTGGCGATT
FVE_5_GW_GSP2	CAGTCGTAGAGAATGGGGACGAGAGT
FVE_exon1_F	AGAGCTTTTGGCGATTCTTG
FVE_exon6_R	CTGAGCTTCAACATCCCAAA
FVE_exon4_F	CAAGATCTCCCTTTGTGAAGAAGT
FVE_exon8_F	AAATGCAGAATTTGCTCTTGC
FVE_exon15_R	GCTCCAGCTCGTTAATCACC
Isolation of a LsFY gene	
FY_ESTseq_2	CAAAAACCGCCCATATCATC
FY_intron1_F	GCATTGTGGTGTTCTGTTTTG
FY_exon6_F	CCAATCAGGTCTATGGTATGGA
FY_exon6_R	GAGCCTCCATCATCACCAGT
FY_exon13_F	TGAGTGGGAGTTTTGATGGA
FY_exon13_R	ACTCCTCATGGAATGGATGC
FY_exon16_F	TTCCATCATGCATTTCAACC
FY_exon16_R	GAAGCGGTTGCATATTTGGT
FY_exon18_R	ACTGAGGTGTTTGTCCATGTG
FY_5_GSP2_3	ATTTGTTGCGGCTGGTGGAATTGATGT
FY_5_GSP1_3	TCTGGAGGACCACCAAGAGTTGTGGAA
FY_5_GSP2	GGCGTCTGCTCACTTACGGGAAAATGT
FY_5_GSP1	CAGTGAGTTATGGCATGGCTGGAAGACA
FY_5_GW_GSP1	GGTGGTGGGTGTTGCTGATACATTGGT
FY_5_GW_GSP2	GTTGTTGCTGTTGTTGCTGAGGATCATTG

Isolation of a LsFCA gene

FCA_ESTseq_v2	AGCTTCAACAGCAACTGCAA
FCA_EST1_F	TGCTTTATTAATATGCCACATCAGA
FCA_EST1_R2	ACGCTCCCCATCAGCATA
FCA_5_GW_GSP1	AGCCAACAAACAATTTGAACTCCACTG
FCA_5_GW_GSP2	ATTGTGTAAGGCACGTATTGCCCTGT
FCA_5_GSP1_2	CCCAAAGCAACATCTTCCATGAACCAA
FCA_5_GSP2_2	TTACATTTGCTTCCCTTCAGCCCCTCA
FCA_exon5_F	GGTCCCATCCAGGTCAGATA
FCA_exon9_R	TGTTTCCCCAGGTCTAGGTC
FCA_exon9_F	GCCATTAAT TGTTCGATTTGC
FCA_exon13_R	TCGCCTGTAAAGTTTGTTTTTG
FCA_exon13_F	CTGTCCACCAGTCAGCTTCA
FCA_exon20_R	GCCTGGACCTGGTGTTTTAT

Isolation of a LsFPA gene

FPA_ESTseq	ATGGGTGTCGGTGTTCCTAA
FPA_ESTseq2	CCATGGGAATATGAACCTGAA
FPA_EST_F	TCTGTGGGTAAGTGGCATCA
FPA_EST_R	TTGAGATCTTAGATAGTCGACACGA
FPA_exon1_F	CCCCGACGTTACAGATTCAG
FPA_exon2_R	TTGAGATCTTAGATAGTCGACACGA
FPA_exon2_F	TCTGTGGGTAAGTGGCATCA
FPA_exon4_R	TCGTCAATCTCACCGAATAAAA
FPA_exon4_F	AGGTGGAACTCCTGTTTGTCA
FPA_exon5_R	CCGGCAAGAAGAAAACGATA
FPA_5_GSP1	TTGAAGGGACTCTTTCGCCCTTTTTGC
FPA_5_GSP2	AACGCTATCAACGGCACCGTGTTTCTC

Towards isolation of a LsCO gene

CO-BBOX_F1	GGATGTGTGATGTTTGTAGC
CO-BBOX_R1	CGTATGCCTTCCTCGAAGC
CO_CCT_2F	GAAGCCAGGGTTTTAAGATACA
CO_CCT_2R	TCCTTTAATTCTGGGTCTTRTYT

LsCO Real time PCR

CO_CCT_3F	AAAGGCTAATCAGATCATTCGCAGA
CO_CCT_3R	GAGAAGAAGCATGTGCATGTGGTG
LsELFα_F_RT	GGAATGGTGAAGATGATGCCGACTAA
LsELFα_R_RT	GCCTCATATCCCTCACAGCAAACCT

Towards isolation of a LsCRY2 gene

CRY2_exon3_R	AAAGCGTATCCCAAAAATACTTC
CRY2_ESTseq	GCGGGTATGAGAGAGCTTTG
CRY2_exon3-4F	GTGCATTCCATGGATGATG

Towards isolation of a LsFKF1 gene

FKF1_ESTseq	GGGCACACACTTTCCTCACT
FKF1_EST_F	CACCCAGAGATGTTGCTTCA
FKF1_EST_R	AGCTGCACATGCGCTAAAAT

Towards isolation of a LsFLC gene

FLC I-BOX–K-BOX_F	GCTGATGATCTTAAAGCCTTGG
FLC I-BOX–K-BOX_R	CTCAAGGTGTTCCTCCAGTTG

Chapter 5 (sequences only written in full if not mentioned before)

Light cycler

FT_EST1_F1 / FT_EST1_R1

FCA_exon9_F	GCCATTAATTGTTCGATTTGC x
FCA_exon13_R	TCGCCTGTAAAGTTTGTTTTTG

FVE_exon4-6_F	CAAGATCTCCCTTTGTGAAGAAGT x
FVE_exon7-8_R	TTGAATACTCCACAAAACAACTGA
FLD_ESTseq	TTCGCGATAAATGTCCACTC x
FLD_exon2_R2	TCCCCTACGCTTTCTGCTAA

FT TILLING

FT_TILLING_F1	GGAGTTGTTGCCAGAGTCGT
FT_TILLING_F2	CTAGGACTTGGAGCATCAGGA
FT_TILLING_F3	TCGTATGATGCCTAGGGAGAG
FT_TILLING_R1	GGTCGCTGGTATATCGGTCA
FT_TILLING_R2	CTCAATGGTCAACCATCATTT
FT_TILLING_R3	CGTACTTCACCCGTAGTTTTCTTATT
FT_REG3_TILLING_F	TGATACCATGTTACCTCCGAAAACA
FT_REG3_TILLING_R	TCGTGATAATATGGTGTGCAAGCTA

FT Real Time

LsFT_RealTime_F	CTGGATGGCGTCAGAACTTCAATAC
LsFT_RealTime_R	TTATCTTCTTCGCCCACCAAACC

Sequencing Primers for PCR products

FLK_GW_seq_exon5_F / FLK_intron5_F / FLKseqR_intron2_2 / FLK_intron2seqR

FLK_EST1_F	TGTTGATTTCTGCAAAAGAGGA
FLK_Extreme_F	ACATTTTTGATGACGTAAGTGAAACC
FLK_5UTR_F	CATCGACAAATCCCTCGTT
FLK_exon1_F	CCGAGAATGAGCAGGATGAT
FLK_F1	CCTCCTGGCACTACTGAAAGA
FLK_TILLING_ F3	AACAAAAACCAATAAAGACTAGGATCT
FLK_TILLING_ F4	CGTCTCCTCAAGGTGGTTTC
FLK_exon2_R	GCACCCTGTTTCCCTATCAA
FLK_TILLING_ R3	TTTGTGACTATGGATTGTTGTGC
FLK_exon6_R	CATACACGGGTCCACCATACT
FLK_Extreme_R	GGCGTAGTTTGTGAGTTGTATTTTCTT

FLD_EST1_F/ FLD_ESTseq / FLD_exon1_F / FLD_exon2_R

FLD_Extreme_F	TTTCAACCAAAATAACTTCAAAAAGCA
FLD_5UTR_F	TCTAGTCACACGTTGCAATTTTTAG
FLD_5UTR_R	TTACATGCAAAATTGAAACAAACC
FLD_EST1_R	CAAAATTGATGTATCCATGTGAGAG
FLD_3UTR_R	GCTTCAAGGTTCCTTTTTCATAAAT
FLD_Extreme_R	CCAATTTGGTCATTCCACTTACACTAC

LD_5UTR_F / LD_RegI-II_F / LD_ESTseq / LD_BACseq_exon7

LD_intron1_F	GGGTTGCATCTGTACATACATCACA
LD_exon3-4_F	TTTGGTTTAACAGCAACACAGGTTT
LD_seq_intron11	CATCTAAAGCTCATGTGCATCC
LD_5UTR_R	ACAATGGAATGCCATCAACATAAAA
LD_EST_seq	CAACTGGTTGCTAGTATTTCTTTGC
LD_RegI-II_R2	GCACTGAAGGACTTTGTATCTGC
LD_RegI-II_R	CATCAATTTGACTTGTCCAGAAA
LD_intron8_R	TTTTTAGCAAGAATAAACAGCTTCG
LD_fulllength_R	AAAACGATAACCAGAAAATGACTT

FVE_exon4-6_F / FVE_exon7-8_F / FVE_exon8_F / FVE_exon7-8_R

FVE_Extreme_F	TTTCTTTAATTCTTTTGTTGTCCTCGT
FVE_comp_F	GGAGAAAATGAAGGAGAGAGG
FVE_F1	TCGTCCCCATTCTCTACGAC
FVE_exon2-4_F	GAGCAAGCTACTTACAAGAATCG
FVE_intron3_F	GCATCATGCTTTCAAACTTCTCT
FVE_seqF2	AAAGTTTGCACCTGGCTTATTGTT
FVE_seqR1	GGAGTTCATTGTTACTACTACTCAATTGCT

FVE_seqR2	TTTAACCTGGACGTTAGCCAAGTAG
FVE_intron3_R	CCCAATTATAGCAATGTCATCCAA
FVE_exon2-4_R	GAATAATGGTTTTAAACTTCTTCACAA
FVE_exon4-6_R	CTGAGCTTCAACATCCCAAA
FVE_seqR3	ACACCTTATTGTGATCCCAGATGTT
FVE_Extreme_R	GTTGAACAAGTGAGAATGTGGGACTT
FVE_comp_R	TCAAGGGGTTGAACAAGTGA

FY_EST1_F / FY_exon13_R

FY_intron1_PCR_F	GCATTGTGGTGTTCTGTTTTG
FY_inton3_F	GCTGCAGCAATCGATGTTAG
FY_intron4_F	ATACAGGGACCAAGTTGAATAGGAAAC
FY_exon6_F	CCAATCAGGTCTATGGTATGGA
FY_intron6_F	CCTGAAAAGGGTTTGAAGGA
FY_Fragment2_F1	TTCTTGTCTTCTGCCATGTCAGCTT
FY_Fragment2_F2	TGGAATCAAAATGGTAATTGGGTGT
FY_intron11_F	TTATCCCTTATGACTGAATAAGCA
FY_Fragment3_F1	CACAAGTGGAAATTCAAAATGCACA
FY_exon16_F	TTTTTCAGGTTTCAACGAACAAAAT
FY_Region4_F1	CCGGAAATTTACAGTCTCCAATCAT
FY_EST1_R	GGACCCCTCTGAAACTCTCC
FY_Fragment1_R1	TTGCCATTTCTAGTGTTTTGCCAAG
FY_Fragment1_R2	TGGACTAAATGCACTCTGACTCTGAA
FY_Fragment1_R3	GCAAAGCTGGTTGATGGATTATCTG
FT_exon6_R	GAGCCTCCATCATCACCAGT
FY_Fragment2_R2	CACAGATTCCTTCAAACCCTTTTCA
FY_Fragment2_R1	TAAAGACTTGGTAGGGTGCCAGTCA
FY_exon16_R	GGTTTTTGGTCTGATGATGTGTCTA

FCA_exon13_R

FCA_seq1F	TTAGAAGTGCTGCTTCACACACACG
FCA_seq2F	ACCTTGTTTTCAACCTCCACAACAG
FCA_seq3F	TCTCTCATTCACCTTCCACTGGTTC
FCA_seq4F	TCTGGAAATCGTCGTCCTTTTGATA
FCA_seq5F	GAGAACAGGGCAACAACAAGGTATG
FCA_intron3_F	TGAGTTTTGTGGTAACTTGATGA
FCA_intron3	TGCAGTTTTGTGGTAACTTGATGA
FCA_exon5_F	GGTCCCATCCAGGTCAGATA
FCA_exon9_F	GCCATTAATTGTTCGATTTGC
FCA_exon13_F	CTGTCCACCAGTCAGCTTCA
FCA_ESTseqv2_F	AGCTTCAACAGCAACTGCAA
FCA_ESTseq_F	GCTCATGCCATCCTTGAAAT
FCA_intron12_F	TTTTTGGTTGCAGTATCAAAATG
FCA_seq1R	CAGGGGATTATGTGCCATTAGTGAT
FCA_seq2R	GAATGTGATGAAACCGGTTATTGGA
FCA_seq3R	AAATTAAACTTCAGCAGCGGAGAGC
FCA_seq4R	TCTTGGGAACAGAACCAACAAAAAG
FCA_seq5R	ATGCAACCACCATATTTCCAACATC
FCA_intron3_R	CATAATGATGCAAATGGAGCAAA
FCA_EST1_R	GCAATGTGTACTGATTGTGTAAGG
FCA_exon9_R	TGTTTCCCCAGGTCTAGGTC
FCA_intron12_R	CCAAAAAGATTTGAGGCTTTTGTAA
FCA_intron14_R	GCATCCCCTGTTAAATCAAAAGAT

FPA_exon1_F / FPA_exon4_F / FPA_ESTseq2 / FPA_EST1_R / FPA_exon4_R / FPA_exon5_R

FPA_fulllength_F	CTTCCCCCATTTGGAGTACA
FPA_exon1_F2	CCTCCTGTCAGACCGAACA
FPA_exon1_F3	CGCTGAAGATTGATTTTGC
FPA_intron1_F	AATAAAAAGTGATGGGGTTTCA
FPA_intron1_F1	GGCTCACCTTTTCTCTGTGTTATTAGC

FPA_intron1_F2	TGGAAGTAATTTTTCAAACCCTCTGTG
FPA_intron3_F1	TGTGAAGGAATCTGAATCTGATGGAAT
FPA_exon4_F2	TCGTTCACATTGACGAACAAA
FPA_exon1_R	CGGTTTGGCAAAATCAATCT
FPA_intron1_R2	ATTTACCCCAGTAAGATAAGCACCT
FPA_intron1_R	ACAGTATAAAGGAAACTGGAAAATG
FPA_exon2_R2	TGATGCCACTTACCCAAAGA
FPA_exon4_R2	CCAGTCTCTAATAGGAACACAACG
FPA_exon5_R2	ACCACCAGCACCACTACCTT

Expression of target genes in LB lines
FLK_exon2_F (CTGCCATTGATGGTTTGTTG) x FLK_EST1_R
FCA_exon4_F (TGCTTTATTAAATATGCCACATCA) x FCA_EST1_R2
LD_BACseq_exon7_F x LD_expression_R (TCGACTGGGACCTTGAGTTT)
FPA_exon_F2 x FPA_exon1_R
FVE_exon4-6_F x FVE_exon4-6_R
LsFT_RealTime_F x LsFT_RealTime_R
FLD_exon1_F x FLD_exon2_R2
FY_exon13_R x FY_exon14_R (GAGGTCCCACACACAGCTATC)

Transcriptome sequencing

164_COL4_F	AAGCCGTGTGATTCGTGTAAGACC
164_COL4_R	GAGAAAGCGGATTAGCCGAATGTAT
164_GI_F	TGCGTGTGAGCTACAGATATTTCCA
164_GI_R	GTTCAGAGATATGAGCAGCGACCAT
164_CIP8_F	TCCATGAAAATCACATCTCCACTGA
164_CIP8_R	CTTAAAGCCCATGAAATCCTCCTTG
Ls_GRP1_F	ATGGGCCACTACTGACCAATCACT
Ls_GRP1_R	TGAGCCTCGTTGACGGTTATGTTA

Retrotransposon iPCR

Tnt77	GTTGACTTGGTTTGGTTGGTAGC
Tnt540	ATCTCGGTGCTACATTATTATTC

Retrotransposon genome walking

GW_RET_5_GSP1	TCCAAGTCAGCCCAATCCTCAGCTTTC
GW_RET_5_GSP2	CAGTGAACGAGCAGAACCTGTGCTCTGA
GW_RET_3_GSP1	TGAGAATCCCGCAGATATGCTGACCAA
GW_RET_3_GSP2	CTGGATGAATGAGACTGGAGGGGGAGA

APPENDIX III

Table A2 - Material making up CGP EST libraries

Species	Library Name	Tissues Harvested
L. sativa cv. Salinas	**QG(ABCDI)**	Calli: grown on MS media (Murashige and Skoog) with 0.1 mg/L IAA
		Roots
		Pre-fertilised flowers
		Post-fertilised flowers
		Fully expanded leaves, seedling shoots and roots: 1.2 mM benzothiadiazole foliar application, 4 hrs and 48 hrs
		Fully expanded leaves, seedling shoots and roots: 0.5 mM jasmonic acid foliar application, 4 hrs and 48 hrs
		Fully expanded leaves, seedling shoots and roots: 50 ppm ethylene for 4 hrs and 12 hrs
		Fully expanded leaves, seedling shoots and roots: 100 uM indoleacetic acid foliar application, 4 hrs and 12 hrs
		Fully expanded leaves, seedling shoots and roots: 100 uM giberellic acid foliar application, 4 hrs and 12 hrs
		Fully expanded leaves, seedling shoots and roots: 100 uM zeatin foliar application, 4 hrs and 12 hrs
		Salt-treated leaves, roots, and pre-anthesis flowers: 250 mM NaCl, 48 hrs
		Drought-treated leaves, roots, and pre-anthesis flowers: water withheld for 48 hrs until wilting
		Heat-treated leaves, roots, and pre-anthesis flowers: 2 hrs at 40oC, 2 hrs recovery
		Cold-treated leaves, roots, and pre-anthesis flowers: 24 hrs at 0oC, 24 hrs recovery
		Leaves and seedlings dark-grown in 0.25x nutrient solution, 14 days
		Leaves wounded with pintool, 2 hrs and 24 hrs
		Seeds: imbibed at 20oC, in water or 100 uM ABA, light or dark, for 17 hrs
		Seeds: imbibed at 30oC, in water or 100 uM GA(4+7), light or dark, for 18 or 24 hrs
		Seeds: imbibed at 30oC, in 10 mM ACC, light or dark, for 18 or 44 hrs
L. serriola	**QG(EFGHJ)**	Calli: grown on MS media (Murashige and Skoog) with 0.1 mg/L IAA
		Roots
		Pre-fertilised flowers
		Post-fertilised flowers
		Fully expanded leaves, seedling shoots and roots: 1.2 mM benzothiadiazole foliar application, 4 hrs and 48 hrs
		Fully expanded leaves, seedling shoots and roots: 0.5 mM jasmonic acid foliar application, 4 hrs and 48 hrs

		Fully expanded leaves, seedling shoots and roots: 50 ppm ethylene for 4 hrs and 12 hrs
		Fully expanded leaves, seedling shoots and roots: 100 uM indoleacetic acid foliar application, 4 hrs and 12 hrs
		Fully expanded leaves, seedling shoots and roots: 100 uM giberellic acid foliar application, 4 hrs and 12 hrs
		Fully expanded leaves, seedling shoots and roots: 100 uM zeatin foliar application, 4 hrs and 12 hrs
		Salt-treated leaves, roots, and pre-anthesis flowers: 250 mM NaCl, 48 hrs
		Drought-treated leaves, roots, and pre-anthesis flowers: water withheld for 48 hrs until wilting
		Heat-treated leaves, roots, and pre-anthesis flowers: 2 hrs at 40oC, 2 hrs recovery
		Cold-treated leaves, roots, and pre-anthesis flowers: 24 hrs at 0oC, 24 hrs recovery
		Leaves and seedlings dark-grown in 0.25x nutrient solution, 14 days
		Leaves wounded with pintool, 2 hrs and 24 hrs
		Seeds: imbibed at 20oC, in water or 100 uM ABA, light or dark, for 17 hrs
		Seeds: imbibed at 30oC, in water or 100 uM GA(4+7), light or dark, for 18 or 24 hrs
		Seeds: imbibed at 30oC, in 10 mM ACC, light or dark, for 18 or 44 hrs
L. sativa cv. Salinas	**CLS(XYZ)**	Seeds: imbibed in water, light or dark conditions, for 24 hrs
		Roots (excluding tap root) from four-weeks-old plants, hydroponically grown in Hoagland media
		Etiolated seedlings: seeds imbibed in water at 20oC for 12 hrs in light, transferred to dark for 69 hrs
		Shoot apical meristems: dissected from plants at 5-6 true leaf stage (approx. 6 weeks old)
		Defence-induced leaves: harvested 24 hrs after treatment
		Defence-induced seedlings: harvested 30 mins after treatment
		Flower/seed development series: 2-8 mm long buds, at anthesis and 6 days post-anthesis flowers
L. sativa cv. Salinas	**CLS(LMS)**	Seeds: imbibed in water, light or dark conditions, for 24 hrs
		Roots (excluding tap root) from four-weeks-old plants, hydroponically grown in Hoagland media
		Etiolated seedlings: seeds imbibed in water at 20oC for 12 hrs in light, transferred to dark for 69 hrs
		Shoot apical meristems: dissected from plants at 5-6 true leaf stage (approx. 6 weeks old)
		Defence-induced leaves: harvested 24 hrs after treatment

		Defence-induced seedlings: harvested 30 mins after treatment
		Flower/seed development series: 2-8 mm long buds, at anthesis and 6 days post-anthesis flowers
L. serriola	**CLR(XYZ)**	Seeds: imbibed in water, light or dark conditions, for 24 hrs
		roots (excluding tap root) from four-weeks-old plants, hydroponically grown in Hoagland media
		Etiolated seedlings: seeds imbibed in water at 20oC for 12 hrs in light, transferred to dark for 69 hrs
		Shoot apical meristems: dissected from plants at 5-6 true leaf stage (approx. 6 weeks old)
		Defence-induced leaves: harvested 24 hrs after treatment
		Defence-induced seedlings: harvested 30 mins after treatment
		Flower/seed development series: 2-8 mm long buds, at anthesis and 6 days post-anthesis flowers
L. saligna	**CLL(XYZ)**	Seeds: imbibed in water, light or dark conditions, for 24 hrs
		Roots (excluding tap root) from four-weeks-old plants, hydroponically grown in Hoagland media
		Etiolated seedlings: seeds imbibed in water at 20oC for 12 hrs in light, transferred to dark for 69 hrs
		Shoot apical meristems: dissected from plants at 5-6 true leaf stage (approx. 6 weeks old)
		Defence-induced leaves: harvested 24 hrs after treatment
		Defence-induced seedlings: harvested 30 mins after treatment
		Flower/seed development series: 2-8 mm long buds, at anthesis and 6 days post-anthesis flowers
L virosa	**CLV(XYZ)**	Seeds: imbibed in water, light or dark conditions, for 24 hrs
		Roots (excluding tap root) from four-weeks-old plants, hydroponically grown in Hoagland media
		Etiolated seedlings: seeds imbibed in water at 20oC for 12 hrs in light, transferred to dark for 69 hrs
		Shoot apical meristems: dissected from plants at 5-6 true leaf stage (approx. 6 weeks old)
		Defence-induced leaves: harvested 24 hrs after treatment
		Defence-induced seedlings: harvested 30 mins after treatment
		flower/seed development series: 2-8 mm long buds, at anthesis and 6 days post-anthesis flowers

L. perennis	**CLP(XYZ)**	Seeds: imbibed in water, light or dark conditions, for 24 hrs
		Roots (excluding tap root) from four-weeks-old plants, hydroponically grown in Hoagland media
		Etiolated seedlings: seeds imbibed in water at 20oC for 12 hrs in light, transferred to dark for 69 hrs
		Shoot apical meristems: dissected from plants at 5-6 true leaf stage (approx. 6 weeks old)
		Defence-induced leaves: harvested 24 hrs after treatment
		Defence-induced seedlings: harvested 30 mins after treatment

APPENDIX IV

Table A3 - BAC clones ordered from AGI

Probe	Filter	Field	Replicate	Address	Plate	Probe	Filter	Field	Replicate	Address	Plate
FT	D-C	3	8	**M22**	**141**	***FKF1***	D-B	2	1	**F2**	**50**
							D-D	2	3	**H6**	**158**
LD	D-E	2	3	**I8**	**206**		D-D	3	8	**M14**	**189**
							D-D	3	8	**N16**	**189**
FLD	D-A	1	2	**G12**	**7**						
	D-B	4	6	**I17**	**82**	***FLK***	D-B	1	6	**D19**	**79**
	D-C	1	1	**O7**	**97**		D-C	4	2	**H23**	**106**
	D-C	3	5	**I14**	**123**		D-C	4	7	**B5**	**136**
	D-C	5	1	**C23**	**101**						
	D-D	1	2	**F13**	**151**	***CO***	D-A	2	2	**B14**	**8**
							D-A	4	5	**M4**	**28**
FY	D-A	2	2	**B24**	**8**		D-A	5	2	**I13**	**11**
	D-B	4	8	**D1**	**94**		D-A	6	5	**B23**	**30**
	D-D	1	3	**I13**	**157**		D-A	6	8	**F10**	**48**
	D-E	2	1	**G5**	**194**		D-D	2	6	**G13**	**176**
CRY2	D-A	1	7	**H5**	**37**	***FVE***	D-A	6	6	**H17**	**36**
	D-A	5	5	**I8**	**29**		D-A	5	8	**J2**	**47**
	D-C	6	4	**A17**	**120**		D-B	4	6	**D19**	**82**
	D-D	1	4	**J11**	**163**		D-B	4	2	**B13**	**58**
							D-B	1	3	**A10**	**61**
FCA	D-A	2	4	**P17**	**20**		D-D	6	3	**B10**	**162**
	D-A	4	5	**D12**	**28**		D-D	6	1	**H9**	**150**
	D-B	1	1	**J8**	**49**						

D-B	1	1	**I10**	**49**	***FPA***	D-A	5	7	**K17**	**41**
D-B	3	2	**M23**	**57**		D-B	2	1	**O8**	**50**
D-C	2	7	**A18**	**134**		D-B	5	8	**F7**	**95**
D-C	6	6	**I10**	**132**		D-D	2	8	**G16**	**188**
D-D	2	3	**B21**	**158**						
D-D	4	3	**B7**	**160**	***FLC***	D-A	1	8	**C15**	**43**
D-D	1	6	**N10**	**175**		D-A	5	8	**A16**	**47**
D-D	6	7	**F5**	**186**		D-B	6	7	**B7**	**90**
D-D	3	3	**F4**	**159**		D-B	6	8	**K9**	**96**
D-D	5	5	**M1**	**173**		D-C	2	7	**N6**	**134**
D-E	1	1	**P24**	**193**		D-D	1	8	**F3**	**187**
D-E	1	1	**I14**	**193**						

APPENDIX V

Table A4 - Plant species used in *FT* phylogenetic analysis

Species	Common name	Abbreviation	Accession/Locus ID
O. sativa	Rice	Os_FTL8	Os01g10590
O. sativa	Rice	Os_FTL1	Os01g11940
O. sativa	Rice	Os_FTL9	Os01g54490
O. sativa	Rice	Os_FTL5	Os0239064
O. sativa	Rice	Os_FTL13	Os0213830
O. sativa	Rice	Os_FTL6	Os04g41130
O. sativa	Rice	Os_FTL10	Os05g41180
O. sativa	Rice	Os_FTL3(RFT)	Os06g06300
O. sativa	Rice	Os_FTL2(Hd3a)	Os01g06320
O. sativa	Rice	Os_FTL12	Os01g35940
O. sativa	Rice	Os_FTL4	Os09g33850
O. sativa	Rice	Os_FTL11	Os11g11870
O. sativa	Rice	Os_FTL7	Os12g13030
O. sativa	Rice	Os_MFT1	Os01g02120
O. sativa	Rice	Os_MFT2	Os06g30370
H. vulgare	Barley	Hv_FT1	DQ100327
H. vulgare	Barley	Hv_FT2	DQ297407
H. vulgare	Barley	Hv_FT3	DQ411319
H. vulgare	Barley	Hv_FT4	DQ411320
H. vulgare	Barley	Hv_FT5	EF012202
A. thaliana	Thale cress	At_FT	AT1G65480
A. thaliana	Thale cress	At_TSF	AT4G20370
A. thaliana	Thale cress	At_MFT	AT1G18100
A. majus	Snapdragon	Am_FT	AJ803471
T. monoccum	Wheat	Tm_VRN3	DQ890163
B. napus	Oilseed rape	Bn_FT	FJ848917
B. oleracea	Cabbage	Bo_FT	EU984306
P. persica	Peach	Pp_FT	EU939302
V. vinifera	Grape	Vv_FT	GU133629

APPENDIX VI

Full length sequences of lettuce flowering time gene homologues

FT

Coding sequence (exons) highlighted in red

The polymorphism at 1101 bp identified in Larissa line 199 is highlighted in yellow

The polymorphism at 1393 bp identified in Larissa line 185 is highlighted in yellow

AGAAATGTGGTGGTGGACTGGTGGCTGTATTCCAGCAGTAGGACTAATTTTCCGGCCAATTTCCGGCAGACAACA
ACACATTCCGACAAATCTTCAACAATTCCGGCAAGGGACTAGGGGCTGAAACAATCCGACACCACTAACGAGTA
TGTCACATAGAGAGAAGATAAAGAATAAGAAAAACATAAAAGAAGGAGTTGTTGCCAGAGTCGTTATAGAGAG
AAAGGGTCCACAAAAGAGGGAGAATGAGCTGAAGGGGGTTATATCTACGAAGGGAATTAGCGGTGACGACTTTT
AGCGGCGACAGATAGTGTATTAACGGCGACATATAGAGTATTAGCGGCGACATGTTTTGGAGAAAATTTTGATG
GTCAAATTTTTGAGTATTATTGATTTTCGGTCATTAGATTAGGGACAAAATCAACGGGTGAGGTTTCACCTCGTAT
GTCGACGCGGATCAGGGTCGATAGGAATTAGCGGCGACAGCCACCTTTAACAGCGACACATAACGCGAAAGTTA
CATGTCGTCGTTAAAGATGGTTGTCGCCGCTAATTCTCATTATTCTTGTAGTGTACTTCTGACTTAATATTTTTGGA
ACCACAAGTGACACTAACCTTAATACTTGTGAGATTCCAATTAGAGACCTAGTTAATTACTACTACTTAAGATAC
TCATAATACGAGGATAAAACTTGAAAAATTCTGTCTCTCGGCGCCTCTGTCCCTTAAATTTTGTTATACATGTCAT
CTAGTGATTTGAGGTGCCGGATCAATGACGTTTATTTATGTGATAAGATTACAAAAATGCAATCCGATTCCAAAT
GATAATGATAACTCTATCTATCTATCCAGATTACTTTTACCTTCGACTTCTAGAACTAACTAGGTGTTGAATTCTT
GTTTCTGAATCTAAAGATAGAGTCCAACATGCTTTGGTAACCATATACTGCGTGAAAGTGCAACCCTTTTGACAA
GTCTATCACTCAGCCTTCGCCACTGGTCTATCTATCATAGGAGAATCCTCAATATTTTAAATCGATCACATCATAA
AGGGAATATGTTTGTGTGTGATATACTTGTGGTGCCACACACTATGTTCTCCAACTCCAAAAAGAAAAATATAAA
CCCTATAAATACCAACACTACTTCCATGCAAACCACTACAACATCCAACCGGCTTTAAAGAAACACAACAAACTA
AACCCAAATTCCAATTTTCGTATGATGCCTAGGGAGAGGGACCCGTTGGTTGTTGGACGTGTGATAGGAGATGTT
CTCGACAGTTTCACAAAGTCGATTAACCTTTCTGTAACATACAACGATAGAGAAGTTAGCAATGGGTGCGAGCTA
AAACCCTCTCAGGTGGTAAATCAGCCTAGGGTTGATATTGGAGGTGACGACCTCCGGGCTTTTCACACTTTAGTA
CGTTTAGTTTCTTTTCATTCTTAATCCTCCTTACTTCTTTCACTTTTTCTTTCTTCCAAATTCGCTCTTGTGAACTAT
AATTAAACTATTTTAGTTACATTATTACTTTTTTAGGTGATGGTCGATCCTGATGCTCCAAGTCCTAGTGACCCTA
ACCTTAGGGAATACTTGCATTGGTAAGCTCCTTGTATAAGAAGACATTTATTATGTGTGAAACGAAAAAAAAATT
CAATTTATGTTTTTGGAAACCATTTTAGCAAATAGATTAGGTTTTTCAAAATGTGACAATATAGTGATAGACAAT
GGATTGCTCCTTTAATAGTGAGGGTTAAACCGAAGTAATAATTTCTTTAATTTTATCGGTTTAACCTAAATTATAA
AGGGCTTGGATCTAAACCCTAATCAATAACGTGACAAAATCAAAGAAATTATAACCATGATGAGCAACTTTAAA
GAGAATAAAAAGTCTAGGAGCATATAAGTGGCTTTCATGGCTTTAATCTGGCTACCACATAAACAAAAAGCAAG
CTTAGATATACAAGGTTCTCACTTCATTATTAAACTCAATGGTCAACCATCATTTATCACTATTACTTTCTAGCTTT
ATTTTTGATACCATGTTACCTCCGAAAACAACATATATTTGTTTTTATTGTAAAAGTCGAATCTCAGATACTCTTTT
CTACATCATATTTTATGTTACTTCTGATTGGTCTACTCGTTTGCTTTTTGTAATACTTTATTCTCTGCATGTAAACC
ACATAGGTTGGTGACCGATATACCAGCGACCACGGGAGCACGTTTTGGTGAGTGCTTTGTTTTTTTTATTTCACAT
CCAACAATTCATATTGTTCGTTGCTCCACTAGTTCTGAAATCATTTATTGCGGCTAATTACATAAAAATTGGAAAA
AAAAAATCGCGTAAATAAAAACTATACTGTGGTCCACGAACTCTATCCTTCCACGATTTTATGCTGATCGCTACTT
CGTTTTATTTTTCTTAATAAGTATATATGGTAAAAGGTGAGCAAAGGAAGAGGTATTTTCGCTAATTTACATAAC
AATCCTCAAAGGTAGGGTGGTTTTGTTTTTCATAAGAATCCAAATTTTATTTTGAGCATACGTGTTCTTGCAAACA
AGGGTACTATAAAATATTACTTTTCTATTTTTTATTATTATTATAATAATTATTATGAATCAACGTCACTGTCGTCT
GTACAACTAAAGTGAAATCTATAAGCAATATGCACTTGAAAGGATAGGAATTTCATAAAGCATTCACTCTTTTAA
GTCATCCAAGAATCTTTTCTGCCCAAAGCAAAACAAAAAGATATTCAAATAGACACTTAATCAGTCTCTAAATGA
TGTCAAATCTTAAATTTTTGATACTAAAGCACGTGTGAATGTTTATTCTCGTTATAATCGATTCCTTAGAACATGT
TGAAAGCCATAACAAGATGGGGTAAGGGTCAATCTCCCACAAATTCGTCTCTAGTGTGCTATAAACTTTAATTAA
TGTGATTGGATTCCCGGATCAAATGCCGAAAGCAACACCTTAAGTTCATTTATTTTTGTTTGATATCTTTCTTGATT
GTTTTTTTGTAATGTAGGGTTTAATTTCATTCTTATTAACTTTTTTTAGGCCAAGAAATTGTGTGCTATGAGAGTCC
AAGGCCGTCAATGGGAATTCATCGCATGGTTTTTGTGTTATTCCGACAGTTGGGTCGACAAACTGTGTATGCCCCT
GGATGGCGTCAGAACTTCAATACTAAAGACTTTGCGGAGCTTTACAACCTTGGTTCTCCGGTGGCAGCAGTCTAC
TTCAACTGCCAACGTGAAAGTGGGTTTGGTGGGCGAAGAAGATAAGATAAAAGCATGTTGATGACATTCAATGA
CGTTATAAGTGCATATGGATAGCTTGCACACCATATTATCACGAGTCTTTCTATATGATATATGTTATATGTTATA
TATCAAAAGAATGTAACTAGGAACCATAGTCATAATTAAAGGAATAAGAAAACTACGGGTGAAGTACGTAGTTG
TAAGATGTAATATAATTAGAAACGATAGAAAAATGATGAGTTGGAATGAAATAAATATACATCATTTTCATTATT
GTTTATTGTTACTTGTAAATACTAAGTATTTTGGATTACATAGCTCGAAAGAATGGAATGGATCGGATACCTAAT
GCCAAATAAATATTAAACATTAGCTAAATTTGCATAATTAACATGTTTCAACAAATGTCAAGATGTTTTTTTATGA
CCTGGTTTTATTCACTAAAAGAAATACGAGTATTAGCGGTAACAACCAACATTCGCGTTAGTTGTTGTCGCTAAA
GGCATTGCCGCTAATACTATTTGACCATGATCCGAGTGAAAGTCCAAGATTACACCCTACCCATTGATCTTGTTGC
AAATCCAACGGCTTGAACGTCGTTTCATATACCAATAGTGGCGACACCAATAGCGGCGACGGTGTCACCGCTAAT
ACTCTAAACAGATGGCGTCAAACTTCCCTCCACTCATGTGTCGGCACTATTGACATTTTGTGGCCACTAATACTCT
AGACGAATAATACTTCCATTTTCTTTTGTCTTGTGTTATGTGTTCCCTTCTCATCTCAACACTAATTCTCTCTGCAC
CGAACGAATAATCATCAATATATTCGATTTCGCCAGAAAAAACCCCAACCCCACACATCACCGGCAACTCCCATC
ACCGGCGACAGGTAACGATGTCTAACTATGTACCGTTTTTTTTTAATTTCTTTTACTTTACTCTTTTCCCCACTGAA
AACTCGAATTTGTAGTGATTCTTTACTGGAATTTCCAGCTTTAGTGTCAGAATTTAGAGCTCCTACCGGAAAAAAT
TTGGAACATCACTTATTTGTCGGAATTGTGGAAACCTCGCTGCAAACTTCGAAATTACACCACCACACGATCCTTT
TCATCTTTTTATCTTTCATTTTGTATGTGGAATGTTTATATTGTGTATATATGATATGTATTTGTATGTGTATGATGT
ATGTGTTTTTCCATGTGTATTTGTATGTGTAGTATTTGTGTATGTGTATGACTTTTGAAATTATATACGTAATTTGT
ATGTATGGAAAATGAAAAAAAAGTAGGTTGTTATAATGAAAAAAAATATGATGCTATAATGAATATGTAATATG
TATAGGAAGTATGTTTAAGTGTATTTGTATGTATAATGTATGTATTTATGTATGTAATGTGTGGACAAAATGAATA
TATATAT

ATAATGCTATAATTGAAAAAAAAAGTTTGTGTATTTGTATGCAACTAGAAAAGTACAACGCTATAATTGAAAAA
AAAGTGTGTATTTGTATGCAACTTGGAAAGCACAATGTTATAATGGAAAAAATGAAAAAAAAAAGTTTGTACTTT
ACAAACTATTTTTTATTTTTATATTTTTTCCATTATAGCA

FLK

Coding sequence (exons) highlighted in red

GGATCACATTTTTGATGACGTAAGTGAAACCGGATCGGGTTCGNATCTGNATCGGATCAGATTCGGGGTTTTGTG
CATCTCTAGTTACCACTACCAGTGAAAATGGTGAATGATCGCGGCAAGTCAAAAGTTTGTGGGGTCATATTTGAT
AAAACCAAATATGCAGGGTGCAAAGAGAAATTATCCAAACATGAATCCCTATTCTCTCCTCCATAAACTACGATA
AAAGCCATTCGATTCTCCACAACGCGCAGCCAAGCCGTCTCTATACCGCTCATTGAATTACGAAATCATCGACAA
TCGACTTCACTTTAAACCCTGGTAAGTTTGTTATCTTACTACTCTGATTTTACGTACGCTTGTTGATTTGTATATGG
ATCTGTATGGGGTTCTTGACGAAAATTTATGTAATTGCGTAGTTGCCTAGTTTTTCTCATCTTCCCGTGGAATCTTT
TGGTGAATTGATTTGTTTTTTTAGATTAAAACACGTAGCCATTTGTCGGCTGTTGAAATTTGATAGGTTGGGTTAA
TTTTCCGTCTAGGGTTTTGTACAGATGCATGCAATTGTATTGAAGAAGTTTCATCGACAAATCCCTCGTTTCAAGC
ATTTCAATTTAGGGATTTTAGTCAGTATACGTTTCTAGTTGCGGAAGGTGCTGTAGGATATGCATTCGGTTGCTAA
TTTACAGCCACCTAATTTCAAAATGTACCACCTTTTAATTAACTTTTTTTAGCTATAAAATCTATAATAAATAAGT
TTAGAATACACAACAATTCTTACATTTGCCAATGCCAAATTGTAGCTGCGTATGTTATTTCTTTTGAAGTTTTTCCT
ACAACCTTAGAATAATAACCTTTTTTGTTCTGCTAACTGTCTATGGCTTGTTCTTAAGTCCTTAGTGGGTGCGCAT
ATTGATTGTGTGATTTTACTTCTTTAAATTAGGGTTTGGGTCCTTATTTCTCTGATGTTTATATAAGGTTTATGGTG
TTATTTTTGGGTTTGGATATATGAGATGAGCTTACTTGATTTATCCATTTCAATCTTTTAGAAAATGGCAGAGGAT
AACATCAGTGAACATGAGACAGAGGCAATGCCTGAGAACTCTGTTCCAGTTGAAATTCAAGATCCCGAGAATGA
GCAGGATGATGCTGTAGAAACTCAAGATAATGAAGATAATGAGGAAGAACAAGGTGGTGCTGATGTGGAGGAT
GCTGACGTGGCAGGTACTGCTGTAGTGGACACTGTTGCAGAAGGTGCTGTGGAAAAGAAATGGCCAGGGTGGCC
TGGAGAGAATGTTTTCAGGCTGTTGGTTCCTGTGCAAAAAGTTGGTGGCATCATTGGTCGGAAGGGAGAGTACAT
AAAGAAAACCTGTGAAGAAACAAGGGCCCGCATTAAAATTCTTGATGGGCCTCCTGGCACTACTGAAAGAACTG
TAAGTTTTTTTTTTTCAAGAAATTTTTTTCATTTTGTTTCTCTCAAAATATAAATTCAACTATCATGTAGAATTTAA
GGTAATGGAGTTGCATAGTTGTGTTTAATATTTTAGATGCCTATATACTTCTTGAAACAATCTCTTATAGTGTGTG
TGTGGTATATGTCTCTGGTTCTGGAATTTGACTAGATAGAGTTGATTTGAAGTGTGATATTGGTTGTTGTTTTAAT
GCAATCTGATATGGACCTATTAGGAACAAATAAATAAAAGTAAACCAACACACAATGTCAAATACTTGAGAACC
AAAAAAGGGAAGGGGTTTGTTAAAACTTAAAAGAATGTATCATAAATAAATGCAAGATGATATATATAATTATC
AAAGAAACTGGAATGTTAAGATAAAGATGTAAAATAAATGTAACTAACAACCAAAGAGAAAACATGAGAAGAG
TGAAATTACAATTTACAAACAAGAATTAAGAATATGAAGAGTGAAGAATGAGAGGTTAAATCTAAATAAGTTTA
TGTTNTCCCAATAACTTAGATGAATGTGTGCAGGTTCTCTCAAATATAAGGAACCATTTCGATTCCAATTAGAAG
ACAACCAAAGATCCAGATTGCAGTTCAACATTGACATTGAGCATCAAAATCAAAAGTTACTTTGATTTGATTGGC
TGGCAGGTAACTTTGGAACTGGATGGATTGATGTCTTGCCACTTGTTGTGCTAGTCAAACAAGTAAACAAAAACC
AATAAAGACTAGGATCTTATATAACATAAAATCAATTAAAAACTAAACAGATATTAAAAAAAAAGCAAAATATG
ATCCACATAAGTTAAGAAACAAGAACATCTTAAAATAGGTAATAACACACTAAAAAGTTAGTAAGGTATATCAT
TAATGGTTGCACAATATGCAGCACTCAAGAATTTACAAGATAAATCTTTGGTATATAGTACAAATCAGAACCTCA
TTATTAGGTGTAGTTGCAAAGTGGTGGCTAGGGCTGAAAGTAACTAACTTTTTTTTTTTTTTTTTTCTTTCTCTATTA
CTATCTACCATCCTTTTCATTGCTTCTTTCTATATGTAGTAATCCAATCATTATACGATTATTGTTGGAAAGTTTAA
TAAGGTGGTTTCTTGATAGGTGTTGATTTCTGCAAAAGAGGAACCAGATGCTTTAATTCCACCTGCCATTGATGGT
TTTTGAAGGTTCATCAACGTGTGATGGATGCTGACAGTGATCCTGCACGTGTCCAGCCTGCCACCATCTCCACTA
GGCTTCTTGTTGCAGGGACACAAGGTGGAAGCTTGATAGGGAAACAGGGTGCCACCATTAAAACCATTCAAGAT
TCTTCAAATTGTAAAATCCGTGTCATAGGAGGTGATTGATTGATTGATTATATTTATAATTATTTCTTTTGTTGGG
ATGATAAATTTTGATGTATGTTTGGAAACATGCAGAAAACCTACCGATTTTTGCTTTACCTGATGATAGTGTTGTG
GAGATACAAGGGGAGGCGACAGGTGTACACAAGGCTGTTGAGCTGGTTGCTACTCATCTAAGAAAGTTCTTAGT
TGATCGCAGTGTTATTGGGATATTCGAGATGCAAGTAAGTTGTTCCCTTGTGATTACAATGGAATCAGCCATTCC
ATTCCATTCCTTTAATTTTGTTTTGAGTTTTTGATAGATGCAAATGCCAAATCCTCGTCAACATCAAGAGATGCCA
CCTCCCCAACCGTGGGGCCCAGGCCCAGGCCCGTCTCCTCAAGGTGGTTTCCCAATGAACGCGGGTGGTTTTGGG
GCCAATCAGTTTATGCCACCTCAGCAGCACCAGTTTGACAACTTTTACCCGCCTCAACCAATGGATAAAGTACAG
CCTCGCCAGCAGGGCCCACCGCCTAGCTTTGGCAGGGATCCATCCATGGGACCACCTCACAGTCACACTTCAACT
CCTCAACCACAGGCACAACAATCCATAGTCACAAAGGTATTATAATAATAATAATAATCATATAAATTATAAAGT
CAAAAAGAAAAATATTGAAGATGGAAATCATAAATGTATATGTAGGTGTCGCATAACATGCAGATCCCACTTTC
ATATGCTGATGCTGTAATTGGTACATCTGGCACAAATATTAGTTATATCAGACGAGCAAGTGGTGCAACAATTGC
AATTCAAGAAACACGTGGAGTTCCTGATGAAATGACTGTTGAGATTAATGGATCTGCTTCACAGGTTCAGACAGC
TCAGCAATTAATTCAGGTAAATAAATAATTAAATACTCCAAAATTTATCCATCTGTTATTGATATTTTTTATTTTTT
GGTGGTTGGAATTTAGAATTTTGTGGCGGAAGCTGCAACTGCTGCTCAGAATTCTGCTGCAAGTGCTCCACCACC
TGCCCAAGCATATAATCCGTACCCTGGTCATGCTCCTCCTCCGGCTTACGCGTCTCAGCCACCCCCCACTGGCCAT
GCACCTCCTGCTGCCGAGTATGGTGGACCCGTGTATGGAGGGAATTATGGGTATTGATTGAAATGTCCTTCCTTG
TGTGAGATTGGTTTCTGTTATGAATGAATGGGAACATGTTGTAACAGGATGGATATTCTTGTTTAGAAATTAGGT
AGATTGTTTGTTTCTTTGTTTGATTTGTTAGTGTTGAGAACCTTCGTTTTGAATGATATTTTGAAACCTTAATCCAA
TCAAAGGCGTTTTTTTTTTTTTTGTAGGGACCTCACCTAAACAAGTTAAAATATCAAGCCACCATATATTGGCTAT
TTTTCATTATATTTCAGTTTTGTCAACCAAAGCTACATAGTTTTTTTTTTTCTATGTTAAAGTGGTTGCTATAGATA
TTTTTGTCAGAGTTAATGGCTTAAATTGAATATTAGTGCCAAAAATTTCACTAAAACCATTGACAAAATGTTCTAT
TGTGGTAACCACTTTGATTTATTTGTCATAGCAACGGTTGACAAAACTGAATTATAATGGTTAAACCAGACATGC
CACAACTTCAAAATATTTAATAATAATTATAAATACTGTTTTTTTGCTAAATCAGTCGGTTCTATTTTTTTTTTTG
TCTCAAGTAACTCTACATTTAATTGCTCAAATATGATTATGTTGATCTTAGATAACCAAAAATCAAGATATGAAT
GTTTATTTTCCATTTACTTCTTTCAAACTCGAGTACAAGAAAATACAACTCACAAACTACGCCTAAAGTTAAAAA
ACCTGTCATTCTCAAATGCATGTTGAAAAGTGACATCTTTTCATCTATATTCAATTAACCAAAAGTCTTAACTGTT
TTGAATAAATGATTTATCAGGAGGTTGTAGACTTGTAGCTATTTGAGCAAGAAAAGCATGATAGACTCATAATTG
ATCTTTGCAATTCTGAAACTCATGTTGGAAAAAAAAATGAAATTACAGTGTCNTAATTAAAAAAAAAACAAATC
ATTATATAATAAAAACAATGAATGAGAATATGTAGCTATTTCATGACTGAAATAGCATTTAACCCTGTAAGGAAC
AGGAACTCCCTCAGTTTATATGCTACAAGCCTNCANNTAGTTGTATGTCTTGACCTCTCTCATTCAACATAACTTG
TGACCAAAAC

FLD

Coding sequence (exons) highlighted in red

TNAGCTTCAAAAAAAACCTTTTTNATAAGCATGTAAAACGTGACTTTTTAAAACTGTTTATGGCTTTTCAACCAAA
ATAACTTCAAAAAGCATTTTTTAAGCCAAATATTTTTTAACTTATTGACTTATTGGTTTTTTCATTATAAAAGTTTA
TCCAAACATTTTTTAAAAAACTATTTTTACCATCACTATAAGCAAATAAACACTTCTATCTAAAAACACTTTTAGA
GTTCCAAAAACACCTTCAAACACCCGCTAACTATATTATTCTAATGATTTAGGTCATATGTAACATAATAGCAAA
CAACTCACATAGTTGAGTAATAAAAGAAATATTTGGTTCCTATGTTAATTCTCTATGATGTTTTTGTTTTTTTGTTT
TTTTTTTCTGATTGCTTGCTCAACATAATATTTAAAATGTTGGACTGCTTACCATCAAATTGTAGGCTTTATTCAGT
TTTCACTTTATTTTAGCATTAAAGATTTGTACTTAAGATTAGGGTTTATTTAAGATGAATTTAAAAAAAAAATGGA
TAATAAAAGATATTTTAAGATTTTCTCATTAAATTATGTATCTAGTCACACGTTGCAATTTTTAGATAAAAAAAGG
TATATAATCAATTCATAAAACACCAAATTAAAACAAATGATCTAATTATAATTCGGTTTGTTTCAATTTTGCATGT
AAAATCACATAATATTAAATTAGAGGAATGAGTTTATAATCATAATTGTAACTTTTTTAAATTGGTTAGGACACG
TATAATAATAATTTATTTAAGATAATATATATTTTTATAGTATACTAATATACNTNTATATTTATGCATTTCGACTT
TTCGAGTCTTATTGTTATAATCTAAAACAACATAACTTAATAATATTTTTATTTTGATTGTTATATTGTATCATAAA
TGGATTGATAAATTGATCTTAAGAATTGGTACGTAGTTAACCCTCTTTATAACTTAAATATCGTTTTTATCAAAAA
TAAATAAATGTTAATTTACAATAAGATTCTATGAATATTTTAATATAATTTTTAACAGTAGCCCATTGCAATGTAT
CATCATCGTCCATGGCGCCTGGTATCAGTTCTTGCTGATCAAATATATTCTTTCGATTCAAATTTCCAACCATGAA
CCCCTCGAACGAAACCCATCAACAACAAATCCTCCCTAACGTTCACAACCCTCTGCAATTCACCATCGTTCTACC
CCACTCTGACACCAATTCAAACCCTAATTCTTCACTCTCAAACCCTAGTAATGTCAACGACTTGCTCACACTCGGT
ATTCCTAGAAAACGAAGACGAGNCAGGNCACACGATACTCTCTCCTCCAATCAGGTATACAATTTACCTCCCGAA
TTTCCCAACAATCACCAAACAACACAAAATTCTTTCGGTTCTAATCATTTTGATAGAAACCCTAACATCGGATCA
AATCAAGCAGCACAAGTAGGTGTATCCGATGAAATAATAGTAATCAACAAAGAAGCAACATCCGAAGCTTTGAT
TGCTTTAACCGCCGGGTTTCCGGCGGATGTACTTACAGAAGAAGAAATCGACGCAGGAGTTGTTTCGGTTGTTGG
TGGAATCGAACAGGTCAATTACATACTGATTCGAAACCATATCATCACCAAATGGAGAGAGAACGTATTCAATT
GGGTAACAAAAGAGATGTTTCTCGATGTTGTACCTTCGCATTGTAGTAAGCTTTTAGATTCTGCATATAACTATCT
ACTCTCACATGGATACATCAATTTTGGGGTTTCACCATTGATCAAAGATAAAATTCTGGCTGAAGCTAATACAAA
AGGGAATGTAATAGTCATAGGTGCTGGTCTTTCTGGACTTGCAGCTGCAAGACAACTGATGACATTTGGGTATAA
AGTTACAGTTTTGGAAGGAAGAAAACGTCCAGGTGGACGAGTTTACACCAGGAAAATGGAAGGAGGAAATAAA
GTTGCAGCTGCTGATTTAGGAGGAAGTGTTTTAACAGGTACACATGGAAACCCTCTTGGGATTTTAGCCAGACAA
TTATCATATCCTCTTCATAAAATTCGCGATAAATGTCCACTCTATCGATTAGATGGAACACCAGTAGATTCCAAA
ACTGATACCAAAGTTGAAATAGCAATCAACTCCCTCCTAGACAAAGCAAGCAAGCTTAGACAGCTAATGGGTGA
AGTCTCACAAGATGTTTCTCTAGGGGCAGCTTTGGAAACATTTTGGGAAGTCGACAACAACTCTTCAAATTCAGA
AGAAACAAGTCTGTTCAATTGGCATCTTGCTAATCTAGAATATGCAAACGCGGGTTTACTCTCTAAACTCTCTTTA
GCATTTTGGGATCAAGATGACCCGTATGACATGGGAGGCGACCATTGTTTCTTGCCAGGTGGCAACTGGAGATTG
GTCCATGCGTTAGCCGAAAACGTCCCGATTCTTTATGAAAAAACAGTAAATTCTATTCGTTACAGCAGTGATGGA
GTACAAGTTGCTGTGAATGGAGGACAAGTTTTTGAAGGAGACATTGCTTTATGCACTGTTCCACTTGGGGTTTTA
AAAAGTGGGTTCATTAAATTCACCCCTGCATTACCTCAAAGAAAATTAGATGGAATCAAACGATTAGGATTCGGT
TTATTAAACAAAGTCGCAATGCTTTTTCCTAATCCCTTTTGGGGCACCGATTTGGACATGTTTGGGCATTTATCCG
AGAACCCAAGTACACGTGGCGAATTCTTTCTATTCTACAGCTATGCAAATGTAGCANGTGGGCGCGTGTTAATGG
CTCTAGTGGCAGGTGAAGCTGCACGTTCATTCGAAAGCATGCCACCCACTGACGCGGTGAAACGTATCCTCCAAA
TCCTACGTGGCATATACTCGCCACAAGGGATCGAAGTCCCGGAACCGATACAAACTGTGTGCACACGATGGGGT
TGCGACCCGTTTTCTCTTGGGTCGTATTCAAATGTTGCGGTGGGGTCCACCGGGGACGATTACGACATTTTAGCA
GAAAGCGTAGGGGACGGGAAGCTTTTTTTCGCTGGGGAGGCGACCACGAGACGTTACCCTGCAACCATGCATGG
CGCGTTTTTAAGCGGGTTAAGAGAAGCTGCTAATATTTACCATTTTGACACCAATCGAGCTATCAGGGGTAAAAT
CGAGAAAAAACCCGTCAAAAAAACGCGTATGTTTGCGGGTCGGTTCTTGCGGATCTGTTTCGGGAGCCGGATCTGGA
ATTCGGGAGTTTTTCGGTTATTTATAATAAGAAGGATTTGGGAGCGATGTGTGTTTTGCGTGTGGGGTTTGTGAAA
AAAGGGCAAGAGAGTTGGAAAGAAGGTCAGTTGTATTCCAATAAGTTAATATTCCAACAGTTGCAATCGCATTTT
AATCAGCAACAAGAGGTTCATGTGTATACTTTGTTATCAAGGGAACAAGTGTTGGAACTGAGGGAAGTGAAAGG
AGGTGATGAAATGAGGTTGAATTATTTGATTGAGAAACTTGGAGTGAAATTGATTGGAAGGAAAGGGTTGGGAC
CTGGTGCTGATGATGTTATTTCTCAAATTAAGGCTGAAAGAAGTAAAAGCAAACCTGCAGTTACTGTTAAATCTG
GTGAACTCTTTGATATTTTTTTTTTTAATGTATTTACAATACTGCCACTGGCGGTGATATTCAAGGACTAACATTTT
TGTATTTTATTTATGAAAAAGGAACCTTGAAGCCGAAATCAGCTATGAAGCAGAAGTTGGTGAGGTTGGTTTTTT
GTATTTACATCCAAATGTTATGTTACATTGTTTGATTTTTTTTTTTTTTAATTTTTTAATTTTTTTGTGTTTCANAAAG
GCGAAGATATTGCGGAAAAACAATAATCAGTTGCTTCCTGTTAAAGGGGTTGGGGGTAATGATGTGGAAGTTAA
AGTAGGTGGTAGTAGTGTAAGTGGAATGACCAAATTGGTGGTTAACGAGTCGATTCCAAGTGTAAATATGAATG
NGGGTGAAGGGCTGATTGGNAATTTTAGCGCCTCTGCCNTCTGCCTCTGCCTCTGATCCTCCGAATGNTNGAGGG
ACTTGGTGTTGGGNGTTGGGGNTGCTTCNNNTGATTGGGATTTAGTAGAATTAGATGGTGAAATGTTGTAAAAAA
AAAAAAAAACCGGGACTTGTGTAGATGTTTGAGTTTTAGATAAAGATGTATTTTTTTTTTGTTTACTTTGTAATGG
GTATTTATTTAATTATTTAGATGTCTGTTTGAAATAATGATGTTTCACCTTATTTGTTGGCGTGACCTAATTTGTTG
ACATCATCATTTTCCGAGGGAAGCGCCAACATCAACGGTCAATTGCGTATAAAGAACAAGTAAGGACAAATTAA
TCATAAAAGTGATTTAGGTTATAAAACCCAATGAAACATGAAAAAGATAGACAAAATTTTTAGTTGACAACTCCC
AACTAAAAGATAAGAAATGAAATTTAGATTCTTTGCGTTTATTTCAAAGAATGATAAAATCTGTTTTGTCACAGA
ATTTAATGATGAAAATTTTATATTCCAATTCCTACGGAATCCATGACATGAAAACCACAGGTTTAGGTTTTGGGC
ATTCTCTTTACATAATCTTTCTAAAAAACGATTCAGCCATTGTTTATTACAACAATTTTAAATTTACAAACTTTATT
TATTCATCAATCGTCACTACCTCCTAACCATTCCTCCAGTTTGGCCTCTTTATCTGCTTTAAATTCATCAAATTTTT
CAGACCTTGACAGGTTCAATNTCNNNTNNNCNTTATCTTCATTCCCATTCTCACTTGATACAGGCAATTCTTCCTC
CATNGAGATCCGA

LD

Coding sequence (exons) highlighted in red

The polymorphism at 2022 bp identified in Larissa line 307 is highlighted in green

TTCATATAGAGNGGTNGAAATNAAAAAGANTAAAAANTAGAAATGTTNNNATTNNNTTTTGTAACCACATCATC
CATNTACCCTCTACAAAAACAAAACCTCTAGCAAGGAGNTAAAGAGAAGTACAAATTAAGGTCGGATTGGATTT
GGTAACCACATCACGGATAGATGAAAACACAGGAAGGTCCAAATGAAGCCAATACCCAACCCGACTCCAAACCT
TNTNANCAACCGAAAATACAGCAACAAATACTTCACACTTTCTTGAGAATTTATGNCACACATGACACANNNNT
NTNAANNNNGTCACCACCCCCATTTCCACCGGTTTAACTCNTACCAGAATGCGATCCAAAAGCACACACCATAA
AAATGTTAATTTTAATAGGGTCAAAACGGTTCCACCTAGTAGACATACTACCCATAGGAAGCGTCATATCATTAA
CCCAAATATGTGTTGATTTAACAGTAATGATAGCCGACGATTCAAGTTTCCTCTTCTCTCTCTTCCGTCTTTCTCTT
CTCTCTCTTCTTTCCGATAGTGGTGAGTATACCGGTGGTTCATGGTGCCCTGAGGTGGCTGACTTGGAAGCTCTCC
ACGACCGTTGTCGTGGAGAGTATACTAGATGATTTAATTATCCTTCTAAGTCATTATCCTTTCACCTTGTGGTGGG
TGCTAGACTGCTAGACGGACAAAAGAGACTTCGTGCTCAGAATGCCTTATATTATTTGCAAATTTTTTGTAACCTA
TATTGGTAATTGATTTATAATTGATTTTATTTTATTTTATTTTATTTTTATGTTGATGGCATTCCATTGTACTG
ACCCCCGGATGGTGGGTAACTGGGTATACCCACACTTTCCGTATCCGGCCCCTCGTCTCTCTATCGCGAAGTATGC
ACGTTCCCAAAACAAACCGTTCTCTTCCAAACCCTAACCCTAACAAACTAAAATCCACGATACGTCTCATGATCA
ACATCAATCTTGCACCATAATCCAAATTATTCTCGTAATCATGTGGATCTAGCTCGGTGAATTAATAAGAGATGG
CGATATCGAACGCGAATCATCAGCTCGCTCTGGTAGTACCATCGAACACGACAACGTCGTATGAGGGGCTGTTGG
AATCGCAGAAAGAACTATTGAGCAAACAGATCTGTGAGCTGCAAAACATCGTTTCAAGACAATGCAAACTTACT
GGCGTCAACCCTCTCTCTCAAGAAATGGTGCGTGGGAACTCGAATTTTTGTTCATCTTTTAACAATACTTTATCTG
ATAGGGCTGCAATTTGTTTTAATTAGGGGTTGCATCTGTACATACATCACACATGGGTCAAAAGCACTAAATTAG
TCGTCACTTGTTAGATGAATACTTTATCAAGATATGAATTATAAGTATTTCTATTTTCGGAGTTGCTTATTTATATT
TTCAGATGTCGACTGATGAGTACTGTTAGTGTTAGACTTTAAAGAGGAGAGAAAACCGGTGATTGAATTTAGGAA
TGAACTTGAGATGATGTTTCTGTAGCACAGTGGACAGGGCGGTGGTAATGTGTGACACTTAGTTCAATTTGAGAA
CAAGGTTGCAGGTGAACCAAATTCACCAAATAGATGTATTTGTTTTCTTGCATGGTTAGCCTCTTGATTATACACG
GAGGAGTGCTGTGCAATGGACTTTTACTTTCAAGTAGATCTGTTATGTTCATAAGGGCAGAGTCTGCACTATAGA
TTTCTATACAGTTTGCAACTGCAATGCAATCTTTTAGTTAGCAAAGAAATACTAGCAACCAGTTGTAAACTTATAT
TTACTTTTGTTGTATACATATCCCAAGCACTGAGTTATCTTTTGACAGTATATATTGTATGTGTACTGATCATGTGT
CTTGCGTTATAATTGCAGGCAGCTGGTGCTCTTTCTATTAAAATTGGTAAGCTGCACAACCTTCCTAGATGCTATT
GATTGTGATCCAACATTTTTTTTTTGTAAATTGTTGTCTTCTAAAACTTAAACCTACCTGGTCTTTCTCTAATGTTT
AAAAGTTATTAACTATTATAAAAGGTTGATGTATAGGCTTATAGCATACTACTAGTAGTAAGAGGTCTGCAGGAG
TGTGCAATGACATTTAGTGTGCCTTTCTTTTTGTTGTATTTTCAGGGAAAAGACCCAGGGATCTACTGAACCCAAA
AGCAATAAAATACATGCAATCGATTTTTTCTGTCAAAGATGAAATCAGCAAGAAGGAAATCCGAGCAATTAGTG
CTCTTTTTGGTTTAACAGCAACACAGGTTTCTGTTTAGTTTATAAGTTCCCATTTAAAGTAGTTTGTATCATCCAAA
CCCATAATAATTAATTCTACTTATTGATAATTCTAAGGTGAGGGACTTCTTTACTAGTCAACGCTCAAGAGTTCGG
AGATTTATCCGCTTATCAAGAGAGAAGGCAACCCANTCTTCTGAATGTGTGCAAGATCAAGATGGACCATCGAG
CTCTAATGTGGACACACTTAATCACCCTGTTCCATTGAACTCTGTGGGCCCATCCAGTGTTGAAGCACCTTCATGT
TCCACCCAAGATGAAGTGTTGCCTGATACAGATGACTCGGATAAGTATTTTATAGCCAATATTTTTAGTCTGATGC
GGAAAGAAGAAACCTTTTCTGGACAAGTCAAATTGATGGAATGGATCTTGCAGATACAAAGTCCTTCAGTGCTCT
TTTGGTATTCTTATATTTATCATTCTCTTTATTATCTATTCAGTACAAACTTCATGGTCTAATCATTCTTTCATGAA
ATTACAGGTTTCTAACAAATGGAGGTGTGATGATTCTAGCTTCGTGGCTGAGTCAAGCTGCAATCGAAGAACAAA
CCAGTGTTCTTCATGTCATTCTTAGGGTACATATATTGTCCAAGATCTTGTTTACAATGTAACATTTATTTCAATCT
ACTGATGGAGCCATGTTTAGGTTCTTTGTCATTTGCCATTACATAAAGCTCTCCCAGCACACATGTCAGCCATATT
GCAAAGTGTCAATAAACTTCGGTTTTACCGGGTATCAGGTGTGTTTTCTATAATTATTGTGTTTGTCAACAAAGTG
AGACATTGTTTGTTTGCTCTTTTGTTATTTGAATCTCATTATCTGATTTGAAGACCAGATGTCTCAAACAGGGCAA
AAAGTCTGTTGTCTAGATGGAGCAAAATGTTTGCAAGAAGCCAAGCCATGAGAAAGCCCAATGCCAACATTTCC
GCAGTTGATGCACAAAATGAGATATTGCTGAAGCAAAGGCAAGCGTCTCTCTATGTGTTTTTTTAGGTGCCTAAT
TATTTTCTATGGGTTTTATATTCATTGTTGTTTTATGATCTTGTATAGCATTGGTGAAATCATGGAAAATGAATCTT
TCCACTCAAGAATTGAGAACCCTGTGAGTTCAAGTCTCGAAGCTGTTTATTCTTGCTAAAAAAAAGCTTAGTTAC
ATTGTACAGTTGTTAACCTCTTAAAACATATTTTCTTAAATAGGCAGCGATATATTCATTGCTAGAGAACTCTGAA
AACTCAAGGTATGAAGCATTTTGTATGGATTCAGACTTGAGGGCTTTTCAACTATAAATCTTGATTATCCCATGTC
ATACAGGTCCCAGTCGATCAAATTGCTCTCATCACCATCTGATGACTCAAATATGAAGCTTCTTAAAGGCGTATC
TTCATCTCGTATCCTTGTGGTTATATGCATTTTTCCATGGAGACCAACAAAATTAGCAAAAAATTACCAATTTAAC
TATTTAACTAACTAATTTAACTTTATGTGTGTAACTAAGTTATTGCCTTAACAATGTTTTAGATGGCAGAGAACGC
AGGAAAGTACAGCTGGTGGAGCAGCCCGGCCAAAGAGCAGGTGGGCGGGGCCCACAGGTTACAAGGTTGTTTCC
TACAGCTCAGGGACGTCCACTTTCGGCTGATGATATACAGAAAGCCAAAATGCGGGCTCAGTTCATGCGTAGTAA
GTATGGGGAAAGTTACGTGGGCTCTCACCCTCACGTGAAAACTGAAGTTCCTCGGGTGTTGATGTCTGCCACCTC
ATCATCATCATCGTCATCATCTAAAGCTCATGTGCATCCTAAAGTTGAGGAACATGCTACTTCCTCTTCTAGCATG
GTGACTCCTTTAGCACCTAAGCCTAAGACCCCTCATGTTCAGGTTCAGCCTAAAGTTGATGAGGAAAAGAAGGAG
CTAGTAAAGAATGAAGTGCAGGAGGCGGAGGAACCTGTGTGGAAGAAGTGTAAGAGATTACCCATATCATGGGT
GAATCCACCAGGTATAAAATCCACTTTCAGATTTGTATTTTTTTTTTATATATATGTTTTTTGGGAACTGGCTGTGT
GCAGGTTCAATCGCAAATTCGTTGTGCTTCTTTTTGTTCATGGTCGAAAGACACTTTAAAATAGAACCCAAAAAA
TGGACTAATTATAATGAATAAAGCAAGCCTAAATTACAAAAGTCATTTTCTGGTTATCGTTTTATTATGGATATGG
TACAATAGTTTCATTGAATGCTTTTTTTTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTGTCCCTCTAAATCT
ATAAAGATGGTCATGCTTTAATCCTTGCGCTTTGTGAAATCCATAAATATGACCATTTTACCTTCAAATTTGACCA
AGACCCACAAAATTTAAAACTAAACCATTAATCCAAATCCAAACTAGCAAGGATAGTTCAGTTCTTTACAATACT
ACCTAATTAATATAAGTAAGATTAGTAGTTTAGTGACAAGGGGAAGAACAGTTGAGGTTGAGGCCCAAGGTGGA
AGGAGTTGAAAGCGGGTTGTGCAGCAATTGGGTTGGGGGCAAAATGGTCTAACTAACCCAGGGATTGAAAGCCT
GATCAAGCGAATCATTACGGGACAGAAAAAGAAAGACCATACCAATAAAAAAATAATAAAACAATGGGTTTTTT
TTTTTTTTTTTGTAAATTAATGAATTAGTTTTAAGAGATTGCAGGAGTTGGATATTGTTTGGGTAATTTTGACAAA
TTAAACCTGTTTGACTTTTTGCAATGTTTTCAAACTTTGTCCATGTGACTGTAACCCGTTGGAGATAAAAAGCACC
ACCAAACTAAAGCCAAATTGTTTCTTAGCAAAAATGAATAAATTTTAAACTAATTATTCTTGAATAAAGTAGTTG
CAAGCGCGTTTCTGTTACTGTAGCATATAACCTATAACCGAAGTTAACGAACTGGTATCTCCATCTTCCTGTTCCA
TGGAGGGTACAAAAGAGACTAGCACTTTATAATTCCATGAAACTGCCACAGGATTTTCATTTTATGCCCTATGAA

ACTTTAAAAAATGAGGATATATCCAGGTTTAGTATGATCAAAAAATGTCCAGTGCTAGTCGTTTAGTTATCATTC
ATGTTGTTGAGTGTATTCAAACAATCAGGGTAGGGTGTTTTTAGTTGTGTCGCTGGTAATTCATATGCCGTTTTTG
CTCTTTTTATTTATTTATTATTTTAACAGAAAAAGCATAATATTAGAAATGCCATCACAGTACAAGATTTGTACAA
ACAACCAATTGATAAGACATGCGATTTGTAATAATCAAGGTGTACATAACAAATATTTGTTGTCCAAGTATGTAC
GAGGTTCTTATGTGGGATGGGGTGGTGTTTTCTGATGTTACTCTATTGGCTGTTGAATATACGGTATGGTTTTGTT
AGTTGGTTTTTGATCAAAATACCCTATTAAGATTATCAATTGATCTCAGCTTGTATAATTAGACAACATTATATTC
ACAGTTCTCACTAAGGAACAGTATATTTAATACTTAAAGAATTTAATCTCAGACAATGAGCAGGATGTGAAATTC
ACGTGTGTGTGTGTGCAATGCCCATAAAACGAGTGAGGAAAATAATGGCAAACCCATCTTGAGTGAATTTGTCA
AAAACGTTAGTTGTATAATTATTTCAAAAGTGCTTTCAATTTGCAAGAAAAAGACCTAGGAAAGAAAGCAGGCG
TCACCTGGAGTTTAGAATCCTAAGATCATTCCATTTTTACTTGCAAATTGCAACATGGACCTCTCAAGTTTAGAAT
TACATTTAGTTGGCATGTAAGTGGTACAGTTGTTCCTGCAAACACGATTAAGCAAAAAATTGGCTTAATTCACGT
GGATTGGAAGTCAGTAATTGTTTGTGTCCCAACCCAATCATGCCAGGGGCATATTTGTCGGTGTCNACTTCTGGT
GTAATCCATATCCATGCATAACTGACTGCCCTAGTAAACAGGCTGTTGGGATGGACAAGTGGTCTGTCTCGAGTT
TGATAGTATTCATGAATTTGTTATTTCTTTCATTCTTCAACTGAGCTGCCTGTTAATCACTGTTAGTTGGCCTAAAA
ATGTTTGAGCTGTCTTGTGTATGCTTTCNTTTTCATGNAGAAACAATCCATTGGTTTTATACGATTCAAGAAAAGA
AAAGACTCAAAAAATAATGTTTGTGAGTTCNTTTGTCCTTTTTTTTTTTAATTATCTATGGAAAAAATTAAAATGGA
TAAAACACACAAATTAAAAAAGAAAGAGCAGACAATATAATATTTTTCATTCATCAAGTCTTGAAACCTAGGAA
GAGTTTTGGGAAAGTTGCTCACTATTTATTGTGTACGTATGCAGAAATGAAAATGAATAAGGAGTGGAGTGTGTG
TTATGGAGAGAACAGCAAAGAAGTTGAGGTCCAAAACAAACGAATCAAACGTGAAAAAGAAGTTTTTTACGACA
CAAATCTTGAAATCCCAGCAAATCCCAAAGAACCATGGGACCGAAAAATGGACTACGATGATTCCCTAACACCC
GAAATCCCAATTGAACAGCTTCCGGATGAAGACGATGAGAATATGAATATGGTAACCCAGGAATCCACACAAAC
AGAAGTTACAGTATCTGAAAATAACAAGCTGAGCAGCATCGGTCCGGATCTCGAACTGTTGGCAGTGTTGCTTAA
GAACCCACAGATTGTGTTTGCATTGACGGCTGGTCAAGGTGGGAGCTTGAGCAATGCTCAAATGGTGAAGCTTCT
TGATGCAATGAAGGCCAATGCGGCTTCTGGTGGCTCCATTGATAACCTTGTTAGCGGCTTATTAGTTGAACCACA
CAAAACCAAAACAGAGGAGAAGAGGGTGGTTGAAGTATCTCTACCATCTCCAACTCCATCAACCAACCCTGTAA
CGGTGCGTCAATGTCAATTATATAATAAAAAACATCATCATGTTTTGTACAAATCCATAATGGTTTATATAATTAT
CGACCCATTTTCTCCTTCCTCTCTCCCTCTCTCTCTTTTTTTACCTTACAATCGAGAAAGAATGCTTAATAATAATT
CCGGTGAATTTGTTTAAAAAATAATGGGTTGAAATTAATTGCACAATACTTGATTTGGTGACAAAGAAAATCGAT
GCGTAAATAGTAAATACTAGATAGTAATAAATCATCGGGATTTGATTAAAACAAAAATATAATAAGTTTATGGTT
TGCTGATTGGTTGGTTTTGTTTTGGTTTATCAGAGTGGAACTGGATGGCGGGCGGGGGCGGAAAGTGGTAAAAAT
CAGTTTAGTTCCCGACAAAGTGTGACGGTTAATGGTGACGCTTACCCCATTCCAGGTGTCCATTTTCAAGAAACC
AGTTTGCTGACTCACCAACAATTAAATGCAAATGCTCCAATTACGCATCATCAAAGATTCACCGATGTGGTCCCT
GACCAGAGAAACATTCCGGCAACAAGTAATTTGACCCAACAGCCAACTTCTTCTCACCAGAGAAACATTCATGTT
CATGTTCATGGTCAAGGAGTGAACACGTCTTCTAATGCTTACCTTTTACCCGCTGTGGGTGGGTCGCATGCATGG
GCGGGCCAGGAGAGTCCACATAGTAGTACAAACTACAATGCATATGGTACAAGAGAGGTGGGTCCGGGTCCGGG
TCCTTCGTGGGGGAGAAACAGCAACAGCAACAGTGGAGGATATCATCCTCGTGGGTTTGATTCATGGAGCCCCG
AGGATAGTCCGGTTAGGTCTCGTGAATATGATGAGAAGGAGAGGGAGGATCATAGCCGGAGGAGCTATGGTTAT
AATTACAATAATAATATGTTGAGAATGCAAAATGACCATTTTAGGTTTAGAGACCGTGGTGGTGGTGGTGGTGGC
CGGAATGAAGCTTCAAGGTGGCGGGATAGGAGGAGATAGATCATCTCCGATTATTGTTGGGTTGATGGACGTTTT
CAGTTTTGATTCTAAAAAGAAAAGAAAACAAACAAACAAAATTGACAACTAAAATCTTGTAGCTTCTTTTTTCTT
AATTCTATCACGTCACGCCCAAAACACATGAAATACAAGTTAGACTCGAGAGTGTTTTGGATAAATTACCACTAA
ACCTTCAAAACAAAACCAAAATTAATACTTTTGGTTTAATCTTTATTTTATTTTGTTTTTAAAACAATTGCTATAAC
CAGCTACCAGGGATTTACAGTTTTCACAACAAAACAATATTTTGAATAATATAAAAACCAAACACCTCAATCGGA
TATAAAGACCTTTTCTGATTATATAGAAAAAATAATGTTTTTTTAAAATCTAATCTAAACTTAAACTACGATATAA
TTTTATTATGTCCAGGATGGGTCTTAAACCTTCAACCTAAAAAAAAATGTAGCAAAACCTAATACTTTTAGATAG
AAAGTCTTTTATTATCTAGGATGAGTCTTGAACCTTCAACCTAAAAAAAAAAATATCCCAAACCTAATATATTCA
AAGNCTTTTAAAATTCG

FVE

Coding sequence (exons) highlighted in red

The polymorphism at 3244 bp identified in Larissa line 307 is highlighted in green

TCAAAAACCCGCACATATTCTGCCCACTACTCAACGATTTAGGGCCACCAACTCGTCGAGTTGCTCTTCCAAAGT
GAATAAATAATGAGTTAAAAGCATACCAGGAACCAGGTGTTACATTTCTATTGCCTCTGAAAGTCAGTCTAGATA
GTAAAAAAGACAGAAATCAAAACTCATAAGTTCAATAGATAATTAAATTTCTTTATGCCTTCGGAAAAAAAACA
AGAGCATGAAGGTTTAGAAGACAAAGTAAGAAAAAAAACAAGGATAATGTAGCAGAAATTTTAGCTAAATCAG
ATACAAAGCAGCAACAGATTTTGACTCTTTTTAACTGAACTCGCATTCAACATGTGTATATATATTTGCTTAAATC
AATGAAAAGAAATCAGAAACAAATAAGAAAAAATGAAGTAAAGAAAAACTAGAAAAGAACTCACTAAAGATAT
GAAGAGTGAATAGATTGAGCCATTCAAAAATCAAAAAAATCCAAGTGACATGTCCTTTCTTTAATTCTTTTGTTG
TCCTCGTGCTTCAGATCGGAGGTTGGAACTCAAATTGCGAGATTAAAACTTGAAAAAAACAAACTCAGACAGCA
GGTGATGATCGATTGAGTGCACAACGGGTTTAGAGAAACCGATAGAATAGAAATCAAATATATATTTGATTTAG
AATGCGAAATTAAGAATACAAATCATATACCAATTTACTGATTTAGAACTCGACTCCTTGATCTTCACATACTACT
CAACGATTTTCAAATTTTAAGGGTGTTATCGAATACCAATTTTAACGATTGCAATTCTCAAGAACTTCAATTATTT
GACGGATGTTTCCGGTGAAATTATTCCCTTAGACATTGATTTATTATTTGACACAATGGACTCGACTTAATATTGG
GCCATCATAATATATACTAATACATATTGGATAAAAATTATGCCTCTAGTTTTTTGATGCCCTTGGCCTTTGCCTA
ACTCGCCCATGTATTGGNCCGATCCTACGTTTCATGTTTCTCATTTGCTGACTATGTAAAAAGATGATGTGGTTTT
CACATATGATCCTACTTGGCTAACGTCCAGGTTAAATTGAAACCTTACGTGGATTTGTTCCTCACTTTCCAGTGTC
GCTCTAGTGGTTCATCATCCCAAGGGAACCTTGCTACTTATCCACAATATCGTCTCCCAAATGAAACTTTCATTGT
CCTGCTTACACCATTTTGGGATGATAAATCATTAGAGCAACTTTGAGAATCCATTGCACAAAATCGTGATGCTGT
GAGAAAAGAAAAGAAAAGAAAAAGAAATAAGAGAATGAGTACAAATCTACATAAGATTTTAATTTAACCTAAA
TATTTGCCAACTAAGATTTTAATTTAACCAAATGAGAAACATAAAACACTGAAACGAAGCAAACCAGTCGATTC
GATAAATTCGAAACAATATTGTAAACTTAATGATATAATAGGAAAAAAACACAATATCAAAACTAAAATTTTGT
ATACATTTAATGTCTTTTATAAAATTTCCCAGGTTTTTATAAAAGATGCTCAAATATATTTGGATTTCTCCAAGAT
AACCTATTATATACAAACAAACAAACGTTTGAGAGTTTTGGCAATTGAAAAGCTATAGTTAAAAAGTAAAAGCT

GTCTAGATTTCACCAAAGCTCAAGGGGAAAAGAAAAACATAGGGCAGAGAGAGGAAAAAGAAGGAGAAAATG
AAGGAGAGAGGCGGGAAAGGCGGAGGTCAGCCGTCGGTGAACGAGCGATACTCACAGTGGAAGACTCTCGTCC
CCATTCTCTACGACTGGCTCGCCAACCACAACCTCCTTTGGCCCTCTCTTTCTTGCCGGTTCNNNNNNATATTCCA
ATTTCTAGGGTTTCTTTTTCNCATCTGCATTAAATCCACTATCTCTTCCTCAATTTCATCACTACTATCTGCATTTGT
TTTACTTTAGCAATTGAGTAGTAACAATGAACTCCTTGTACCTACTAATTAGTAATTAGGGCGATTGATTGATCTT
ACATCCTCGTTAAAAGCCAATTTGACAAACGGAGACATGGAAAGTTTGAACTTAATAACCTTCTTGAAACTTATT
CTCAACTCAGTTGCCACCATTATGGTTTCCATTATCTCACTGACCTTTGTGAATAGGAATATGTTTGACTAGTAAA
ATAATATACTATTAACTGAGCTTATTTTGTGCTTTAAACGTGTCATAGACTTTTACATTCGCTATTTTTTGGGGGA
ATAAAGTTACAACATGCAATTACTAGTAACAAGTTGTATGTTCACATCGACTATTTTTTGTTAATAGATGGGGGC
CACTACTCGAGCAAGCTACTTACAAGAATCGCCAAAGACTCTACCTCTCTGAACAGGTATATGATGATGCTTACT
GAAATCTGTTGTTTGTTCTTATTGCATTCTTGAATTTCTTTAGTTTAAAGATATATAGGTTTGCTTTTGTCAGATAC
CAAGTATTATTTTTTATTACAAATACCAAGAAAGTGATTTCTAAAAATAGTTTTCTTGTATTTTATTGTTTTCTCAT
ATCTCTAAACCTTCTTCCAGACTGATGGTAGTGTTCCAAATACACTTGTGATAGCAAATTGTGAAATTGTTAAACC
AAGAGTTGCAGCTGCAGAACACATAGCACAGGTGAGTCCATAATAACTCCTAATTGTAGTTTGCAATATACTTTT
ATTTATTAATGTTTTTTAACATTCTACTTTCACTTCTTTCTATTACAAACATTCTACTTTTAACTGTTTCCTTTGAAA
AATCTACAGTTATAGCATCATGCTTTCAAACTTCTCTTAATCTATGGTACAAATTGGTTTTTATTTGGATGACATT
GCTATAATTGGGTTCATATTTCAGATTGTAACAAGAAGAATATGTTATTTCTTGCATTTAACCCTTCTTTTTTATAT
CTCTTACACAAACACCAACTTGTGATTTCTAATTTTTGTGAATTAGTTTACTGACTACTACTGATGTTGTTGTAGTT
CAACGAAGAAGCAAGATCTCCCTTTGTGAAGAAGTTTAAAACCATTATTCATCCTGGAGAGGTATTCTCTATTTT
CTTTATGATGTTTTTCATAGATTAAGGTTAAATAAATAAATAAATATTTTTTTATTTGACATGTAATTGTATTTATG
AAGGTAAACCGCATCAGGGAGCTTCCTCAAAACAGCAAAATAGTAGCAACACACACAGACAGCCCTGATGTATG
CATTTGTTTTCCATTTTTCATTTTTGGCACTTCATTTTATTAGTTTTTGATGTATTCTTTTAATGAAATATAATTTTA
TTAAACCAGGTCCTAATTTGGGATGTTGAAGCTCAGCCCAATCGCCATGCTATCCTTGGAGCCACAGAGTCGCGT
CCTGATCTGGTGGGGTTCATTGTAGACAATCACAACACAATTAATACATTTTGCACTTCCTTCTTTCATTTTTCTTT
ATTGTGATTTGTGGGTGTTCTTTTTTTTCATCTGGGTCTGTATAATTTTATCTAAAGTTGATTCCTTTTTTACCCTTA
CAAATCTTTCAGATCTTGACTGGTCATCAAGAAAATGCAGAATTTGCTCTTGCTATGTGCCAAAGTGAACCCTTTG
TGTTATCTGGAGGTGATCTATTTCAAAACCATTCTTCAAGAAAATTTCTGGAAATAGAAATACTTTTATTTCTTTT
ATTTATTTACTGAAACTGTTATTTTATTTTATTTTAGGCAAGGATAAGTCAGTTGTTTTGTGGAGTATTCAAGATC
ATATTTCCACTTTAGCAACAGATTCAGGGGCAACAAAGTCTCCTGGATCTAATGGTGGTAAGAGCCCAAAGACTG
AAGATAAAGTTACTGAGAGCCCTAAAATTAAAGCAAGGGGTATCTTTCAAGGGCATCAGGATACTGTTGAGGAT
GTACAATTTTGCCCCTCTAGGTATCATATCATGAAAATGGTTTTTTCTTCTATCTTCTATTATAGTCATGACACTGT
TTGATTTACATTTACTAAGTGGAAGTTTGTTATGTGGAATAGTGCACAAGAATTTTGTAGTGTTGGTGATGATTCT
TGTCTTATCCTTTGGGATGCAAGAACTGGATCCTCTCCAGTTGTTAAGGTCACTTTAATTTTCATTAAGTAATAAA
GTTTGAAAAATTCTTTTGTTTATAAAATGTATTTTTTTAGTTTATTATATGTATGATGGTTTGCAGGTGGAAAAAG
CTCACAATGCTGATCTCCACTGTGTTGATTGGAACCCGATTGATGAAAATTTCATTTTGACTGGGTAAATCATGTC
ACATATGACATATCTCTATATTCTCTTACTTTTTTCTTTTTTAATTTTTTTATTATGATTTTTAATTTAATGTGTATG
TGTATGTAATTTAGGTCTGCTGATAACACTGTTCGTTTGTTTGATCGTCGAAACCTTACTTCCNATGGGATTGGAT
CACCTGTTCACATTTTTACAAACCACAATGCTGCTGTACTTTGTGTACAGGTATTAAATAAATATATTTTAAAAAA
GATGGTTTTACCCTTGTTCACTATTTCTAAATTCCAATTACTTTGTATGTTTCAGTGGTCTCCTGACAAATCATCCA
TCTTTGGAAGCTCAGCAGAAGATGGTGTCTTGAACATCTGGGATCACAATAAGGTGTGTCAACTCAACCAAAAAC
CACTTTTACTTTTATTCCATTTTCCCCTATTCTTTTCTTGCAACAAATTGTTTTTTTTTTTTTTTTTTTTAATAAAATG
TTTGTTTGTTTGTCTACAGGTTGGTGAACGTTCTGGTCCAGCCTCAAAGTTTGCACCTGGCTTATTGTTCAGACAC
TCTGGACATAGGTATATATATTTAATTTTAATTAATGGTTTTGTTGTTTACAGAAATAAAAAAATAGATAAAAATT
GAAACAAGATTATATTGTGCTATGATTGCAGGGATAAGGTGGTTGATTTCCACTGGAACTCGCATGATCCATGGA
CAATTGTCAGTGTGTCGGATGATGGTGAAAGCACAGGTGGCGGTGGAACTCTACAGGTATATATATCACTCATTT
ATTATTATTTTAGTTTTTATTAAAATGTTTTTAGTTATATTGATTAGTTGAAATGTTTGTTATTTGAATGATGATGT
GTGTAGATATGGAGGATGATTGATTTGATTCATAGACCACAGGAAGAGGTGATTAACGAGCTGGAGCAGTTCAA
GTCCCACATTCTCACTTGTTCAACCCCTTGA

FY

Coding sequence (exons) highlighted in red

GAGACACCCTCTAACATAAAATTTTAATATTTTATTAGGAAAATTGCAGTCGTACATACAATAACCTAAAATTTG
AAAACAAGTATAATACATAACATCAAAGAAAAAAAAATCGGATCAAATCATCAAAACAAGGCTTTTGTAACATT
TTCCCGTAAGTGAGCAGACGCCGAGCGTCTTTCATCAAAACTGTCTTGTCTTCCAGCCATGCCATAACTCACTGTA
TAATCGGAGTCCAGTAAACCCTAATACATTATACCTCCTTAACAATCCCAATTTCTCACTAGGGTTCGAATTTTCC
AACTGTTAAAAGATGATGTACAATGATCCTCAGCAACAACAGCAACAACAACAACAACCACCACCAATGTATCA
GCAACACCCACCACCACCGCCTCCGCAACAGCAAANGNNGCAGCAGCACCCACATCACCAGCACCACCACCAAC
AGCCGCCGCCACATCAATTCCACCAGCCGCAACAAATGGGAGAGTTTCAGAGGGGTCCGCCGCAACAGCAGCAT
CAGCAACCGCCACCGCCACCGCAGATGATGCGTCAGCCTTCTGCTTCTTCCACAACTCTTGGTGGTCCTCCAGAG
TATCACCAACAGCTTAATCACCCCCCTTTCGACGGTAAGCCTAATATACCCTACAGCTCAACAAGTTTTGGGGTT
GTTTAGGTTAATTTTCTGTGTATCCTTTGATTTAAAGTTCTACATTTATCCATAGCATAGTAGAAGACTTGAAGCT
TTAGAGATTTGAACACACTTACCCCTTAGAAGACACATTCGTTTCTACTTATAGCCCACCAAAACCTATATCGAGT
AGCAAATGCTCTAACGCAGAAGCATGATACAATGGAAACACATTCTTTTTTAATCTAACAGAGAACTAATAAATG
TGAAAGCATAATACACTACAAATTCTGTGAGTTCACATGGCTAAATGTGAGGTACTATGATCAATGCCTCTAGTT
CTATGTAACAAGATGGATGAACATATAAATATTAGTTGTTAATTTCAGAGAAAGAGCATCAAGATGCAAACATG
GATGTATAACTCTGACCATTGTGAAAGTAAAACTAAGATTTGGAACTAGTGAAGATTTCATTTTTCAGTGGTAGA
CCACATCCTTGAAGTGATTAGGGTTTTTACATGTACAGATCAACAACAAAACAAGGAAAAATTACCCAAGTCTTG
TAATTATGCTTGGCAAAACACTAGAAATGGCAATGTCAACACCTTATGGCTAAAAATTTGATCATTCTTATAGCT
GATTTCTTCTTCTTTTTGAATTAGTTTGAAAGTCAACAGCCCAAATCATTAGTTTGACCTATAAACTAAATGAAAG
CAACTTATTTGTAAAGATGTCAGATCAAAGAGTTTGACCTAAAAAACAAATAGCATTGTGGTGTTCTGTTTTGAT
TCTTATAGGGTATCTTATATGTTATTTAGTGTTTTATTTTTTTTAATATTTTAGCATTTTATTCAGNNGTAATAATCT
AACTTTTTTTTATACCTGGTTAGTTATTTTTCTGGCTTCTACTTTCTTTTAAGACAATGAGGTTAAGTAAATTCAAG
ATTAGTTTTAACTTGAATTTTGCTGTTATGTTTTCNTATTTAGTTGNTGAAAATCAAAATACCATGTCTTTGAGTTT

TATTATCTTAATTAAAAAGTTATTTATTATGGTTGAAATACAACCCATGCAGATAACTTTGCTGCTAAAAGAATG
AGAAAAATTGGACAACGAANAGCTGTTGATTATACAAGTACTGTTGTGCGATATATGCAGGTAAAAATATTGTTG
TAGATAATTTTTTTTTTCTTTTTCTGATTAATATGACTTAATGCACTTAATTTATAAAAAACTTAATGAGTTCAGCT
ATATATATACCTGATTTTTACTTAATGTTAACAAAATGTCCTAGTAAGAAAAATATTTTAAAATTGNTCTAATCAT
GTTATTTTTATACTTTTTAAAATTGTTCAATTTGATAATGTACAATTTCGTATTATAATGGTTATTAAATAAATGGT
AATTTGTTATTGGTCAGTTTTAAATAGTAGGGTAAAATAGNCNTTTATTCAGAGTCAGAGTGCATTTAGTCCAAA
AAATAAACAACATTTGTTCAATACAAATTATACATTCAGGACTTAATAGGAAAAAAAAACATGGCCTAAGAATC
TAAGATGTTATTTCTTTTGANNNNNNGGACTTANTCTGTTTGTTTCTTTTTGACTTAATAAAGTTTTTTATTTTTAT
TTTTATTTTTTTAATGCAGACACGTTTATGGCAACGAGATTCACGTGACAGAACAGTATTACAACCNNCCCCTGCT
GCAGCAATCGATGTTAGTCACCATTTTTTCTTCTCACAACATTCNAGATTTAAAACAATTTATTTATTTCTTTTTAT
TGTTGTTTTTATAAAAAGTGCTTATTANNTTATACAGTATGAAAACTAAGTTCAAAAGAAAGTGTTTGGATATTA
CCTTATTTGGTACCAAAACACAACATGATAATAAGTTGTTTTTTAGACATGTAAAATGACCACACATAACAAGGG
TAAAAAGGTAAATTGGTATGATGATAAAAGAGAAAAACAAAACTGCAATCCATATGGTTCTTTTAGAGTTAAAA
CAGGGGTGAAAGGTGCTTATTTAAGTGAAGAAACACTTATATATTTTGTTTAAGTGAACCTTAAATAATCAATTG
CAAACACCAACTTACAATGAATTATCATGTTTTATCAATTTACAATGAATTATCATATTTATATCTAATTTAAAAA
TATGTTTAAATGTCAGATGTTACCAACAGTTGCCTATTCAGATAATCCATCAACCAGCTTTGCATCCAAGTTTGTT
CATACATCACTCAACAAAAACCGCTGCTCAATCAATCGTGTGCTGGTGAGGGCTTTTTGGTCAAATTCCCATTATT
TTAAGAACAGTTTTTATTTAAATCTCAAATGCATCTTGATTAGTGAAAACTGAAACCATAGATAGAATAAAACTC
ATAATTCAAAAACATTACAAAATTAATATGTAAACTTTTATTAAAGAGAAGTAAAAATTACAAAATTGGTCCCTG
TGGCATACTGTTTCTTTTTTCTTCCAGTTTTGTTCCAAGGGCAAAATGGTCATTTTAAAAATTTGACACCTAATAC
GAATAAAATACATGCATATAAAATGTTTGTAAAAGAAAATATCAGCCCTAAACTGTTATAAACTAATAAAAATTT
AATGGACTTATAAACTGTTACAAACCTTTATTAAAAATAAGTAAAGATTACAAATTTGGTCCCTGTGGTATACTG
TTTTTTTCTTCTCTCAGTTTTGGTCCAAGGGCAAAATGGTAATTTTAAAAATTTGACAGCTAGGCTGTTAACATGG
ACGATACAACCATACAGGGACCAAGTTGAATAGGAAACTTAAAAATAAGGATGAAAATTGCAAAAACCAGCTAT
ACTCGAGGGTTAAAAATGTAATTTGCTAAAAAATTGAATTCTAACTGTGTCCTGACATATGTATATGTTTAAAGT
GGACGCCTACAGGTAGGCGACTTATAACAGGTTCACACAGTGGAGAATTCACTTTATGGAATGGTCAATCTTTCA
ATTTTGAAATGATTCTTCAGGTAATGCAATTTGATTTTCTTTTTTTTATTTTTTATTTTATATAAATTTTTTATTTTG
AGAAATTTGGAATATGTGAATGAATGTATGCAGGCTCATGATCAACCAATCAGGTCTATGGTATGGAGTTACAAT
GACAATTGGATGGTTACTGGTGATGATGGAGGCTCTATAAAGTAAATAAATTTCTTTTTTTCTTTTTAGTTTTAGG
GAAATCTCCGATAAAGTCCCAATATTTTAGGCTAGTTTACGCTTTAATCCCAATTTTTATTTTTTGAACAGAAAAG
TCCTGATTATTTTATATTTTGCACAAATTTACNAAACAGNTTTTAACTAAATTANATAATCGGGACTTTTCTGTTC
AAAAAATAAATGTTTGGAAGGGTTAATCTCATAAAAGACCTTATATTTTGTGCTTTATTCGGATTAAACCTCNTCT
TTATTTTTTTCCTGAAAAAAACCATGTATTTTTTCATTTTTTTATGGAAAAGCCCTCTACTTTGTCTTTCTTTTTCCAG
AAAAACCCTCAACCTTCTNNNNGNTTCCAAATAAACCCTAAACTTTTTTAGTTTTTCTCAAAAAAAACCCTCACAT
TTAACAAGTTTTTTAGGGAAAAAAGGGTTAATCTCAAAAAAGACCTTATATTTTGTGTTTTGTTTTCTGGATTAAA
CCTCAAATTTATCTTTTCCTGAAAAAAATCTTGTATTTTTTCATTTTTCTCAAAAAAACCNTCAACTTTGTCTTTTT
TCCGGAAAAAAACCCTCAACCTTCTAACTGCTTCCNAATAAAACCCTNNNANNNTTTAAGTTTTTCTCGAAAAAAC
CCNCNCATTTTACTANNTTTTTTCGGAAAAAAGNGACTAAAAGAGCCNGANTTGGAAAATATTGNAAAAATGCA
AGGTCTTTTTCAGGAAAAAAATAAAAGTTGAGGTTTAATCCGGAAAAAACACAAAATCTAAGGTTGTTTACGAG
ATTAACCCTATTTGGAATATGCAAAAGATACCTGAAAAGGGTTTGAAGGAATCTGTGTAAAGATGCTTTTTGTGT
TTCTTTTTAAACCTGTGAAAAATTGATATGATATTCATATGGTGAACGTGTGTAGGTATTGGCAAAACAACATGA
ACAATGTAAAGGCCAATAAATCTGCACATAAGGAATCTGTTCGTGATCTAAGGTATAAATAACATTCTACTTCCA
CTTCTTTTTAACAACATTGTATTTTCTTATTAAGGAATTGCACTTTAAAAAATCCACTATTTGTATCCTTATACTTT
TAATTTTTTTTCTCTCTTATGCTGTAATATGAAGAAATGAAAGTAAAAAAGCTATTTCTTGCATTTCCCTTAATCTT
GAATCTTGTTTTGTTTAACAGCTTTTGTAGAACAGATTTAAAGTTCTGTTCTTGCTCTGATGACACAACTGTTAAA
GTTTGGGACTTTGCTCGTTGCCAAGAAGAACGTTCTTTATCCGGTAATGTTCTTGTCTTCTGCCATGTCAGCTTCAT
GGATTAAATTCAAGTTTATGTTTTTAATATTCAAAATAATATTCACTTGCATCTATAATTGTCAATTATAACCATA
GTCTTTTCTAATATATGGCTTTAACTCAGACAGGCCATGGTTGGGATGTGAAGAGTGTTGACTGGCACCCTACCA
AGTCTTTACTAGTTTCTGGTATGCACTTCAATATATATATGTTCTTTTACGGTTTTACCCTCCTCTTTCCCCATGTTG
AGTTGCTAAAGATAAAAATTATTCATGCTTTTGTGGCATTCTTTTCACACAATAAGCAAACTTAAAATGGTGCAA
AAGTGATTGAAGTTTACTCTGTTTGTATTATTTTTTAGGTGGGAAAGATAATCTTGTGAAGCTTTGGGATGCTAAA
TCAGGGAAAGAGCTATGTTCATTGTTAGTGAAAATGTTATTATTTGACTCCTTGTATTTTAACATTTTGTCCATAA
TGGTCCCTCCTTTGGAGTCTTAAAAGTTTCTTCTTTATGTTTGTAGTCATGGGCACAAGAATACTGNTCTTTGTGTC
AAATGNNATCAAAATGGTAATTGGGTGTTAACTGCTTCTAAGGATCAAATAATCAAGGTGCAGCTTCTTATCTTC
CTTTGATACTACATTCTCATACTTTCATATATACATACATACATACATACTAATTTTATTGTGATTTTGTTTTTATT
ATCCCTTATGACTGAATAAGCATTTCTTTATTTGTGCAGCTATATGACATTAGGGCTATGAAAGAGCTGCAATCTT
TTCGTGGTCATCGAAAGGATGTCACTGGTTAGNTNNAAATCTTATGTGGATATTATTATTATTATTATTTTT
TTAATACTTATATTGCGTTGTTTTCTGATTGATAGCTCTGGCTTGGCATCCATTCCATGAGGAGTATTTTGTGAGTG
GGAGTTTTGATGGATCAATCTATCACTGGCTAGTGGGGTAATATGGTCTNTTAGTCTTTAAAGGCCCCTTTTCTTC
TAAACAATCANGGANNNNNNNAAAGTTGAAACATTTATTTATTACTTTTTTTTTTATACACACAGGCATGAAACT
CCACAAGTGGAAATTCAAAATGCACACGATAGCTGTGTGTGGGACCTCGCGTGGCATCCCATTGGGTATCTCCTT
TGCAGGTAAAAATTATTCATAATTAATTACATATGATATTATATAAGAATAAATATTAATAAAATTACAGTTTTTG
TAATAAAAGAAATGAAAATAGAAATTAGGATATGCTGACATGATATAATTTTTGATGCAGTGGAAGCAGTGATA
ATACTACAAAATTTTGGTCTAGAAATAGGCCTGGGGATCCAGCTCGTGATAAATTCAATGTGGGTCATAATCAAG
GTTTGTTAGATTCACCAATTTTCATTATTTTTCTTAAATTGAAAATTTATTTTGTAAAATATTTTTATTGTAATCTTG
TTTTAGTTTTAATGAGCTTTTATTATGGTCAATGTTTTTATTTCACAAACAATCTTATCGAATTATTGTTAACAATT
TTTTGATTTATAAAAGTAAAAATTCAAACAAAATAAAAGCTCATTAGATCCCAAATAAGATTATAATATAAGTCA
TAGTATATATATTTTCTGGTTAGNTTTTAGAAAATAAAATTGTAATTTAATGGAACATTTCAACCTTTTGTGTTGTT
AGAAACTACATAAAAAAATGGATTATATTGTAAATTTTCATTCCATATTATGATAGAATTAAAAAGAACATATAA
ANTTNANCAAAAATAGGAACCACGATTCCCTTATAATGGTCTATTCCATTCCATTAGATAAANCAATCACTGTTG
AGGGATAAGATAAACCCATTTATATTTAAATAATTTTTAATGAAACAATTTCTTTTATAGAAATCTCTTAGTAGTA
GTTAATATAAATTACAATTTTAGTTTCTAAAAACTAGATAATATATATGTTCTGATAATCTTATGTAAGTTTCAGT
TAGCCTTTTTTTATATTTTGTAAACAATCTGATTTAATTATAATTAATAAACTTTTATATACAAAAGTAAAAATTC
AAACACATAAAAAACCCTCATATAATTATAATAAGAACAAATATCTGGTTATTCTAGAATATAAAATATTAATTT
ATTCAAATCGGGATTTTTTTAAAAAAAAAAAGCTAAAAATTCAATGGAAAAAAAAACATCCTCTGGCTATGAAT
AAAATTTAACAAATGAATTAATCAATCACATTCCTTCATGAATCGTCTATTCCTTTCCATCACACTCAAAAGTCAA

AACTAACATAATCTCAAGGGTATTTTGGTCTTTTCAATAAAACTTTTTTCAGGTTTCAACGAACAAAATCCTGCTC
ATGGTGGGCGTATGGCAGTCAATTTTGCGGGACCCGAACCACCAACAACTCCGGGACCATTTGGTGGATTAAACC
GGAATGAAGGAACCATCCCCGGAATCGGAATCGCAATGCCATTATCTCTAGACACATCATCAGACCAAAAACCA
CCCATATCATCNTCAATGCCACCGCCATTGCCGCCAGGTCCCCACCCATCNNNNNNNGTCAATCAACAACAACAA
CAACCACCCCAACACCACCACCACCAACCACCACATCAACAACAAATGACATCCTTACCCCCNNCNNNNANNCA
ACCGCTTCAACATCCTAGCCATATGCCGCTCCTTCCACATCCTCATTTGCCCCGCCCCCCGCCCCAATTGCCGCCA
CATAACATGCCCGGAAATTTACAGTCTCCAATCATGCATGGTCAAATGGTAATTTTTTTAAATTTCTACACTCATC
TAATTTTTATTTGAAACTAAAGGGTTAAATGTAAATTATATCAACTTACTTTTGTTGAATGTTTTATATTAGCATTA
TAGTTTTATTTCTTCTTATTAAAGCATCATATACTTTAAATAAAATATTATTGAATGTTTTCTTACATACTTTGAAA
ACTTTGCCCTATTATTATAATAGTATTATAGCAATAAAAACTGCTTTGTATTTGCATGAAATTGCTATAAATGAGT
ATGACACACAAATAAATGAAGTAAGTTGCTATTATCTTTAAAACAAAAGGGTTAAATGCAAAAAATCACAACAT
ATTTCTTTTGTTTATTTTAGCATTCTAATTTCATTTCTTCTTATTACAGCATTGTACTTTGAAAAATATAGCATCCTA
CTTCCATCATGCATTTCAACCTTTTAAAAAAAAAATACTAATTAAGGCATTGTACTTTGATATTTCTTCTTATTTAC
TCTGAAAATGTTATCCTAATAGAAATGGAAATTTCTTAACTTTTACACGATTGGTCCCTATGGTTTGGGNTANNNN
NNANATTTGGTCCCTAACTTATTTTTTTAACTCGAAAGGTCCCTACTGTTTTTTTTTTTGCGTGCTTGGTCCTTATC
TTACCTAAAAAGACTATTTTGCCCTTGATTTTTTAATTTATTTAAATAAGGNGGGGTGGGGTGGGGTAGGTAAGG
TAAGGGAAGTGGGGCNGGGGGTGTGTTTATTTAAATAAATAAATCAAGGGAAAAATAGGTAAGACAAGG
ACCAAGCGTGCAACAAAAACAAACAGTAGGGATCTTCCGAGTTAAAAAAATAAGTTAGGGACCAAACGTGCAA
ATTACCCCAAACCANNNNNNNCATTCATGTAATTTACTCTAATTTTTTTTGTAGCATTCTACTTTCATTTCCTTTTA
CTACAGCATTGTATTTTGAAAATTCTGCTCAATTTTAGCAATGTCATTCACATACAAAACAATTTTTATTGCTATA
ATTAGTTGGATTTTTCAAACTACATTGTCTTGATAAGAAATTTCTTTAAAGTACAATGCTCTAATTGGTTAGATTT
CTAAAAGTACATTAATGAAAATTCAATGCTCTATTAGAAAGAAATGAAAGTAGGATGCTATATAATACAGAAAA
TGCTCGTGTTGATAATTTATGTAATGTAATCCAGGGAGCAATGAATCAGATGGGTCCCCCGATGCAGCAAGGCCA
TTTCATGGGCATGAATCCTATGCAGCAAGGCCCACCGAGTGGGGGTGGGGGTGGGACCCAGATGTATTCATCAG
GGGCTACATTTAATCGCCCCCCAATGCCAGGGCTAGGTCCTTATCAGGTAACACATTTTCTTTACACCTACTTTCA
TTTGTTTATCAACTATTATTATTTTTTTTTAATTTAATTTATTTTGCAGNNNGGGAATCCAAGTGGTGGCGGAATG
GGAACAGTGCCACCGAATTTCGGTATGCCTCCGCCACTGCCGCCTGGCCCACCGCCACATGGACAAACACCTCAG
TAG

FCA

Coding sequence (exons) highlighted in red

CGAGCGGCCGCCCGGGCAGGTCTGCAGTAATGGTGTTCCAAATCAACTCCATAACTTTGAAAGCCATTCCTAATT
TCATATGAATTAACACCACTTGGCACTGCATACCAAGTCCTGCAAATATAAGCATTTAGGTTATTATAAATGTTAT
AACAGTCAATTAAAAAGCTATATTTATATACGAAATACAAAATATACCTTATATCCTCAACAAATTCCTCAATCT
TTAGAAGTGCTGCTTCACACACACGTCACCTTCCCATTTCTGTGCTTCATTTCTTAAGTTGTAGAACCTTCCCATCA
TGTAAGAGTCTTATTTCTTCAAGCATAGCTAAAAGAGGCTTCTTTATAGCATCTACAATGACTGCATTAAAACACT
CAACCATCCCATTCTCTACAGACTCATATGCCAAATCACTACCAAAAAATGCCCTACACCATGTATGAGGTTCTC
TAGCCATAAGATGATCATAAGCACCTTCACTTAAAGTCTTGATCCTATCATAAGATGATNATNAGCACNTTCNCT
TAAAGTCTTGATCCCTANNCATGTTTATCTTGAAATCACATTTTGTTGTGGATTTTACAGTTGCCCAAAGCATGTT
TATGAGTTGTATCNTACTATACACTTTCCTAAAATTGGCATAAATATGCCTAGCACATTGTCTGTGCTCTATATGG
GGTAGAATTTCCTTTGTTGCTTCAACAAGACCCTAGGAAATGAAAAAAATATATTAGTAAATGTAATTAAAAACT
ATAATAGATGATATGGATTAAATGAACCTGTCATTACCTTGTATTGATCACTAATGGCACATAATCCCCTGCCAA
CATCCAATGAAAGATCTTCACTAACTGGTTCCAAGAACCATTTCCAATTAACCTTGTTTTCAACCTCCACAACAGC
CTATGCTATAGGGTAAATTTAATTATTTGCATCTCTACCTATTGCAGTTAGTAGCTCACCCTTGCATTGTCCCTTCA
AAAAACACCCATCCAACCCTATAACCCTACGCATTCTATCTTCCACCCATCTTTGGTTGCTTTAAAACACACATAC
ATTCTATGAAAAGTTGTAGTCTGATTAGGGTTAATATTAACAAAAATCCTAACTATGCTACCTGGATTGGACCTC
ATCATTTCATGTCCATAATCCCACACTATTTCATAATGGTCACTCAATTTTCCATCAATAAAGGACAATGCCATAC
ACCTTGATCTATAACATTTTGACCATGACACCTTGCAATGGAATTTACTTTGAATTATTGCCTGTATTTCCTTGCAT
TTTAACTTTGGTTTCCTTAATAGTTCTTTCACAAACTGCTTTGCTAACCAAGTTGGATTCATAAGCCTTGAACTACT
ATAATTCCTAACACATGTATGCCTATCATTCATCTTTTTCACCTAGAATGATTTTCAGTACTTATCAATGAAGCAC
AAACCATAAAAGGGCATTTTTTCTGGATCACTAACACACACAACCACTAGTCTAGCACTATCACACTTCTTGAAA
CAAATTTTATAACCCTTATAGATGGAATAATTTGTTAAACATATTTCAGTTTCATGTCTAGATTCATACCTTTCTCC
AATAACCGGTTTCATCACATTCCATTTGACCTTCAGATTATGGATGGAGTATACGATTTCATATTCATCACTAGAT
TCTTCAGTGTTTTTCTCATCATCTACTTCCACTCGAGCATTCTCTCATTCACCTTCCACTGGTTCGTCATCATTGTCT
TCACAAAGAACATTAAGGAATCCATTTTTGCATCTTGAAACTGATTTGGTTTTTACAAAATCATCAACTTGAACGT
CGATTGCGTCAGGCTCAAATANTTCATCCATCGCCAAAGGACACGAGCAAAATCGATCGATCGATTTGTCGGGA
ATAAAGGAGACGAGAAAAAGGGGTTTTGTTTAATAACCCTAAATTAATGAGGTCAACGATACAATCACCATTAA
GATGTCATGTAGACTTTTGTAACCAGCTAAATGAGGGTCACCGGATGATAAAGGCTAAATTGTCTGGTTTATCCA
GTTAATTGAGCTTTTTCATTCAAAAAATTTCTTTTAGGATTTTTTTGAAACTTTTTCCAAAATATAAAAACTTTATG
GTAGATATCCCAAATATTATCAACCAAAGAAGCATATTATCAACCAAAGAAGCAGGAGACAAGAAATAGCAAA
ACTCATGTCAGGGGGTAAAAAATGGAAACAACAATACCTCGAGGAGTTATATATACAAACATGACCAATAAATC
ATGGTCGCCGGCTCTCCGCTGCTGAAGTTTAATTTCCAATCCCAAAATTAATTTCATAATCTTTATCGCAAAATCT
CAAAAATCTAGGGTTAGGGTTTCTATTATCTCTTGAAATAATTTCTGGAGAAGCTTCATTCGTTCATGGACAGACA
TAGAGGAGGAGGAGGAGGAGACAGGTACGGGAATAGTCCTAACGACTCACATCCTTACAGACACCCTAGAGGT
GGAGGCGGTGGAGGTCCACCACCTTCACGATCTTCAGATGATATTCCAATGAATCGTCACCATGGAGGTTTCTCT
GGAAATCGTCGTCCTTTTGATAACAGTCCACCTCGTTATTCTCTTAGCGGTGGTGGTGGTGGCGGCGGTGGAGGT
TTTCACCCGATGGATGGCGACGGTGGTTTCCGGCCTTTAGGTGTGCGTGGTGCTAGGAGTTACAACAATCCTTCTT
CAGAGTTCGAGGTACCTCTTTCTGGGCAGAAAAGGTTTAATCCTGATTACGAGGTGCCTATGTCGGGCCAGAAAC
GTCAATTTCCTTTTCCTGGCCGAGGAGGGCCTTCTCCAGGTAGCTTGCTTAAATTTACTATTTGTTCGTTGATTTTG
CATGTTAAAGTTGTTATTTTGACCTTCATGTGAAGGAGTATCAAGCATTAGCTACAGCCTCATTCGTCCAGCTTAT
ATTGCATTGTTATTCTTTATAGATCGATTTGACGGAGGCAATTTTGAGAAAAGGAACTTCGATGGAGATAATTTT
GATAAAAGAAATGTTGATGGAGGCAATTTTGACAAAAAACCTGTTGAAGGAGGCAGCTTCGCCAAGCTTTTTGTT
GGTTCTGTTCCCAAGACTGCAACAGAAGAAGATGTAAGTTCCGAACTTGTGAAATGGCATTTTTTTTTGTTACTTC
TACCTAAGACTGAAGTATTTTTTTTTTTTCTAGTTTTCTATAGAATGTGGCAATTCCATCTATCCTGCTAGGCTGAT

TAGCCTGCTGTTGCCTTCTGTTTGCATGATGACTAACTTTGGTATAATGTTAATGAATTCTATATATGATCTCAGA
TTCGACCTGTTTTTGAAGAACATGGAAATGTTGTTGAGGTTGCATTGATTAAAGACAAGAGAACAGGGCAACAA
CAAGGTATGTGCAATTACATAGATGCTTTCTGTAATTTGATATGTGTCGATATCTCTTTTCAGGTCCTTTGAGGGG
CTGAAGGGAAGCAAATGTAAAGTTATGTTTCTTCCCAGAATCCCAGTTAGTTGTTTTGGCAATCTTCAGGCCTCTT
ATTATGGTTTGCAAGATCATATCTGTCTAAGAAGAAGAAGAAGATAAAACTAAACCATTAAATATCAGTAACTAT
AACAGTTGGTTCATGGAAGATGTTGCTTTGGGGGCAGAATAATGTGTCATATAAACATTGCAAGCTTCTTTTTAG
ATGAAGTTTTTGGTACCTAAAAAGCTTAGCTAATAGTGTTGAGATACTGACAGACAACTAGGTTATTTGTAGCAT
GTGGGTAGATTGCAGCATTGTAACTTTTGAGGTTTTCTACTTCTTTTTGGCGTTTAGACCCTGAAAAACTAATCTA
TGAATGCTTGTTAGTATAATTTAGTTGTGATGTTGGAAATATGGTGGTTGCATGGATGGGTTGATAGTTGCTCTAC
TAACATCTAGTTTTCTGGTTGGAATTATGTGGTTGCTATAAGCTATAGACTATAGACTATAGTGTATTGTTTGTGT
TTTGAAATCATATTTGGAACATTTAGGCCTTGTAGGTACATTGTGAGTGAATTTCTTAACTTATCAGTTACTTTTTT
GGGCTCNAGAATCTTGTTTGATTTCATTGGTTGAGTTATTCNGCGACACTAGATTGATTGTCTTTAATGGTTGCTA
TTTTATTCTTTGGAATTGCTGAATGTTTCTGTCATTGGGGGATATGTTTTAGTTTTCCATACAATGGATCTTTTCCT
ATTGCAGTTTTGTGGTAACTTGATGATTCTATTGGTTGCTTTGTGTATGATTTTAGTGGGAAATTAGTTGCACTCTT
TTAACCTATAGTGCATGTTTGGTTTAGTTGTATCATTGATGACATCGTATCATTGGAAATGTTTCTTTTAGCTTTAT
TTGATGAATGTTGGATGTGCATATTTTTTTCTCTATGTATGTAGTTTAAGTTGTGGTGTTGTGATAATATGTGATGG
TTTTCCAATAGTTGTATCTTTTTGTATTTGATTCTTGCCACTTTATTCCATTGTTAAAAATAAAAATGTGTGTATGA
TTCAAATAGAATGTGAATGAGTAGTAATGTCTTTCATTGGTTAAACCTCACAAATCTAGGAACTTTATCATGGTTG
TTGATGAGCTTTTGAGAGACTGAATGTACATTTTGCTCCATTTGCATCATTATGTAGTTGCTTTTTTTACTTTTTGT
GGTTGTGAGATATTGCAAAATAATGGACCTAGGTGTGTTTTACTTTTTTTATGGTTGTGGTTGTTGTTTTAGAGTA
GATGGGATGGGAGATTGAATGATGATATAATTGATTGATTCAAATTCAATAATAGCACATAATCCCAATTGTTTT
ATTATGTTTTCTAGTTGAATGGTTGATTGTAGTGTTGTTAATATTACCAGGACAAACGGTGTTCGCCTCTGCAGGA
TGTTGCTTTATTAAATATGCCACATCAGAAGATGCTGACAGGGCAATACGTGCCTTACACAATCAGTACACATTG
CCAGGGGTTAGTCTTTTTTTTTTTTTAAAAAANNNNAAAAAANGGGGNGAAGNCTACTTAAGTCTGAANNGNAN
TTTTTTTTTTACAGGGAATGGGTCCCATCCAGGTCAGATATGCTGATGGGGAGCGTGAACGTTTAGGTAATTTAC
ATGATAGATGCATATGCTTATGTGGTTTGTTTTCCTTTGGGATTTGATCCTGATGTTGTGTTCTTATGTGTTAGGTG
CAGTGGAGTTCAAATTGTTTGTTGGCTCATTAAATAAACAAGCAACAGAGAAAGAAGTTGAACAAGTAAGTTGA
CATCACATATTCATAAATGTTTACTTTCTTATTTGCTTATCTATTTATCTATATATATACATGAAAATAATGTAACA
AAAATTAATATTCAGATTTTTTTGCCTTATGGTCGTGTTGAAGATGTCTATCTCATGCGCGATGAAATGAAACAGA
GTCGTGGTTTGTCCTCGTGTCATTTTTATTAGTATTATTTTCTAAATTTTTTCTGTAATTTGTCTTCATATGATATCA
ATATTTTGTGTAGGATGTGGATTTGTTAAATACTCAAATAGAGATGTGGCAATGGCAGCTATTAATGCTCTTAAT
GGAATTTACACAATGAGAGTAAGCAATTTCATGCTTCAAATTTTGTCTCTAGAGTTTCATCATTTTACCTTTTTTA
ATTTAAATTTTTTTTTTTTTTATAAAAAAAAATCAGGGTTGTGAACAGCCATTAATTGTTCGATTTGCCGACCCTAAA
AGACCTAGACCTGGGGAAACAAGGTTCTCTCTTTTTTATTTATTAATATTATACGTATAAATATTTATTATAATTT
ATATAGTTTTCTTTTCTTTTTCCAGGGGAACTCCTTCATTTGGTGGCCCTGGGTTTGGTCCTCGGATGCAGCCCCCA
GGAATTAGGTATCCATTCCCATCCCATTTGAATTAGGTATATATAATTTACATTTTTGGTCCTTGAGGACAACCGT
TTCTTTCAATCTTGGTCCCGAAGAGTTTTCCTTTGCAAATTTGGTCCTTGTCTGAGGTGCCGTATTTTTAATCACTC
TTCTTGGGGGACCAAAATAGTCATTTTACTTAAAAATCAGCTATATTCGTGGATCAAAAATGTAATTTTTTTTTTC
TTTTTTACAGACCACCGAATCCTGGGGAGGCAGGAATGCATGGTAACGCATGGCATCCGATGGGCCCAGGTGAT
GTTGGTGGTTTGAACAACCAAATGGCAGCAGCTAGATCTGGTGATGCAAAAGTGTTGCAGCTTTCATCATCGCCT
TTAAAAAAAGCATCAAATCAATCACCACAGCCTCCATCATCCAACACCATAAGACCAATTGGTCAACAACAAGT
AGTAGCACCACCCAATTCCAGTGGTCAAACACATTCACATGGCCATGGTTTCAACACAATGTTACCCCAAAATCA
GTATCAGGTGCCACCTGTCCACCAGTCAGCTTCACCCCTTGCTCAGATGCTGACTCAGCAAAAACAAACTTTACA
GGCGAGTTACCAGTCATCTCAACAAGCCTTCTCACAGCTTCAACAGCAACTGCAACTTATCACACCTCAGCAGAA
TCTTCAACCTCCTCCTCCTCCTGCTGCTAGATTGCAACAGCCTAATTCTGCTCCGCTCATGCCATCCTTGAAATGT
AACTGGACTGAGCACACATCGCCTGATGGATATAAGTACTACTACAATAGTACAACTGGCGAAAGCAAGGTAAC
TTTTATTTATTTATTTATTTTTTAAAATTGATTGATATGTTGAAATACATACATACCTACATACAGTGGGAGAAAC
CGGAGGAATTGGCTTTGTTTGAACAACAACAGAAGCCTCAAGAACAAAAGCAACAACAACAACAACAACAACA
ACAACAACAGCAAACTCAGTCTCTTTCACACAGTCAAGGTGGTGTTCCACATCAGCAAACTCCAAGTCTTTCACA
CAATCAAGGACATCAGCAAATGCAAACTCTCACTCAAGGTGTGCCACATCAGCAAACTCAGATTCCACAGTTTCA
GAATCAGATGCAACTGCAAGCACAGGGGCGACACCCTCAACACTTGCAAAATCCCACTCACTCTACACCCGTAA
AATTTTCTTCCGCTATTATTATTATTGTAGACTAAAATACAGTTTTGGTCCCTGACAATTTTGCTCTTACAATGTTC
TATTTTGCAATAAGTATATAATCATTTTACTCTCTAACACTTTACAAAAGCCTCAAATCTTTTTGGACCAAAATTG
GAGAAATCAGCTATGCTTGGGAACCAAAACTGTTATTTCTTCTTTTCTGAAAGGTTTTGTGTAATTTAATTTAAAG
TAAATAGCTGTTTTGGTCCCAAAGGTTTTTTTGTCTTGCTATTTTGGTCCTTATTTGAAGTTTTTTGTAACAAGTAC
CAAGAAAATGACTATAAATTCAGGATGAAATTTGCAAAAACCGTGACCAAAAATGTTACAAGAAGCTTAAATAA
GGACCAAAATTGTAATATTCGGTTGTACCCAGGGACCAAAAATGCAATTTACTTTTTAGTTTATCTTCTGATTAGT
ATATTTGATGATCTTGTTATCCATTTTTTTTTGGTTGCAGTATCAAAATGCTGGGGTTGCAGGCCATCAAAATATT
CAGGTATTGATTATATTTATTTTTATATGATTTATTTTTTGTTTTTTTTTAATTTTAATAAAACTGATTTGTGTTGCA
GGGCTTTGGGTATGGACAAATGCCAGTGGGTGGTGGTTCCATGAATGATTCTGCACAACGCTTTCAGCAGGTAAT
AAAATTCTTTTAAAAAATCNNCNAAAATTCTTATATTTTTAAATTAATCTTTTGATTTAACAGGGGATGCAGGGA
GGAGGAGGTCAGGACTGGATGTGGAAAAATAAAACACCAGGTCCAGGCAGTTAG

FPA

Coding sequence (exons) highlighted in red

GGTCATATTTGACGTGTCCATTGCTAGTATTTGCATGTCGCTCTCCCTTTGTTTCCGAAGCTCTATTTCAGAAGCA
ATTTTCAAAATCTTGTCCATTCTCTCTTTTATATCCTTCATGTCGGTTCCTATTTCTTTCAAATCATCTGAATCCGAC
GCTTTGCCTTTGCCTTTGCTTTTTTGATTTGTCTCTTCCCAATGGTCGGGAAGGTGGTGAAACTGGAACAGGTTCA
TCTTCGTTCAGATCAAACTGAACACGAGCATCTGATGTAGTGTGGTCGGACTCAGATGTTCTACTGCGTTTTGACC
CACTGTCGATATGTTGGTCAGATGTTTTGCTACTCAGCCATATCCTGTTATCTTTCAAAAAATTCCAAACTTCAAG
AAATTTGAAAGCCTTCAGATTTTCTTTTTGATACACTTTCAACGCTTTCTTCAAAATGACTTCATCGCTTTCTCCAC
TTTTCCATTGACGTTTCATGTTATTCCACAAGTCATTAAACAACATGACTTCCTTGTTCATCTTCCCCCATTTGGAG

TACACTTGATGTTTTGAACGGTATTCTCCACGGTTCATTCCTTTATGGAACCTGTGTAAAACGCGAATCCAAAAGC
GGTCATGCTTTTGGTATTTTCCAACCGTCACATCTTCTGAAATATCGATCCAAGATTGTGCCAACACTAACGCTTC
CAAAGGTTCCCAACATCTAGCCTCCGCCTTCTCTTTTCTTTTAGTCGCTGCTCGAGTGAGCTGAGTTTCTGTAACG
AATTCCGGTTCAGAATCAAAAACAACTGTTTCTGGTTGTGTTTGTGTTTGTTGAACAAACTCGGGTTGGGGACAA
AAAATCAAATGAATTAAAATCCGGGTCGGGTTGTGGTAGTGGTGGTGTTTGTGGATGTGGTTGCCCTGGTCGTTG
TGGAAAAGAAGGAAACATACCGCCATAGTGGTAATACGAAGGTGAGAAAATCGGTTGATCAAATGAATGTATTG
GGGTGTTGGTGTTTCTGACTGGTGTTTTTTCTTTGCTAGGTCTTTTGGAAGAAGACATTTTTTCAAAGAAAATTGG
ATGAAATGATTGTATTTATATGTGTTTAAAGGTTGTGTAAATGTAAAAGTGATAACATATGTATATATAGATGGA
GTAAAAGGTTAAAAAAAAAAATTAAATTACACGAATCTTTGAACGGTCAAAAGAAAAAGGAATAGCCAATCAG
CAAACGCGGTCCTCAAAGGCCGCAGTGCGGTCTCAACCGCCCCAAACCGCACCCTAGGGACGGTGTTCCTGTCCC
ACGTGGCGCAACCACGGTTCCAAACCGCATGCCACACAGCATCTTTTTTTTATTTTTTTATTTTTTTATTTTTTATT
TTAACCCATTATTGAGGTCCAAACCCGCATTGTACGGTCGTATTTAATCCTCGGTCCGGTCCAATCGGTTGTACAA
AATCGAGTGTTTGGGTTTCGCCTCTGCGTTTCCTACTTCTCCCCCCGATACTCATCCAAACCAAAATCTTCAACCT
GATTGTTCGAAACTTCAATCTAACGAATGGGGCCTCCTGTCAGACCGAACAAGCAGTCCTTCGGTACTGCTAATG
CTATTGGGAAGCAAAACCTTTTCGAATCATCATCGGATGCCATGTTTAACAATCTATGGGTTGGCAATTTGGCCG
CCGACGTTACAGATTCAGATTTGAGGAATCTATTTGAGAAACACGGTGCCGTTGATAGCGTTACCTGTTATCCGT
CAAGGAGCTACGCCTTTGTCAACATGAAGCGTCCGGAGGATGCAAAAAGGGCGAAAGAGTCCCTTCAAGGCTTT
GTTTTACGTGGGAGCGCGCTGAAGATTGATTTTGCCAAACCGGTTTGTATCGCCTCATTAGGTTTTCGTGGTTGTT
CTTTCTTATTAGTTTTTCAAGGAATGTTATGATATACGGATCCAAATTTTGTATATGTTTTGTTGATATATAGTTCG
CACACAACAATTATCGTCTTTCAAAATCTCTTAATAACGCTAAATTTGAACAGGGAAGGGGAAATTTAACCGCGA
TGTTAGGGGATTCGTCTAGGGTTTCATGAAAAGAGTTTTTTGTTTTTTCTCAAAAACAAACCATCAATATGAAATT
TTTATACGGTTTAAGATACAAAATAGGGAATTAGAGAACACATTCAAACACTAGTGAAATAATTTGATGTAGTAT
CAGTTGAACAACACGATTTATAGGCTATAGCCAAGTAAATCAGGACAATATTTCCTTTGTTTGCTATCTGACATTC
TAAGCAGATCATATATTCATATTATTTACAATCTGGCGCTATTAGTTTTACCCAATTTCTCAAATTTACCACATAC
TCTGTTGCACGATTTTGATAGAACTGGTGTTGGGATAGGATGCCCTCACCAGTTTTATTCCCCCCCTTGTGATCGA
TCAACTCGATCAATAGTAAACCAATAACATTATGTTACTATTTACTAATCCTGGTAAAACTAATACAAAGAACTC
CCATGGATATAGATATCTAGCAGCTGTCTTTTGGCTTTTAGAACATATGTGCGATGATGATGGAATAACATGGTTT
TCTTTACCTTACTAAAACTTCCCACCTAAGTAATAAAAAGTGATGGGGTTTCAAGTCTTGTTAAGTGCAAAGATG
TAATCTACTATCTGTCAAAGGGTTCTGCTAAGCAAGCGGATGCATTATTTGACATCTAAACAAGTTCTTTTTCATA
CAATAATGGANGCATAACTGCAATACAAGGTGGACAGTTTAGTTCTAAATGTTTAACTTGCAATATAAATCACTT
AAGTATCCTTACAAATCCCTGGCTAAAGAGTAGATATTAATCACCTATTTTCTTCTTAAAGTCAAGCAAATAAAG
TTGTGTGTTTGATTTTTCTTAAAACTTAAAAGAATCAACAAAAGAGTTCTTCTCTAGAAGGTGCTTATCTTACTGG
GGTAAATGTTTAAACTATTTCTTTTTTTAAANNCAAGTATTAAAATNACACANGGAGAACTCCAGTTTCTGGATC
CCATATCTCATAACACATGGATATCTTGTTAGGGCCGCTTAATAGGACACATAGGCTTCCAGCCAAATATAAAAT
AAGGGGGATTAGTTGGGTTTAAACTTGTATCTACATGAAGAAACTATGCTACCTAATATTCATTAGCCATTCATG
AAATTCAATTCAGGTCAATTTGAAAAGTGTTCTGCTCAAACTGAAATACTCTTCTTTGTTTTTTTTACCCAGTTGCT
TTTAGCCTTATTTAAAACAAGCAAACATTGTTATACACTTGATCACAAACTAGGTCCTTCGATTATTTTTCTAATC
CTATAGAAACTTAAAAAGTTGAGGGTATTAGAAAATCAATTTACAACCTTCTTCTGGGTTGACTGGGATCTGGAA
TTATGGATGAAATTGCAAATTGCAAGTCGAATTTACCTCCCAATTGAATTGTGCTCGTTGTTTTACTTGTCTCAAC
TTGGAAGTAATTTTTCAAACCCTCTGTGTAATTGTACTGATGTAAGAGGAAAATTAGTCATTTTAAAAAGTATAG
TTTAAAAGTAATTACCTTTTTATATATCAAGTTATTCAAATAAACTGTGTAATGTGTCAGCAATACACCCCGCTTA
TCTAGAAGGTCACTCAACTGCCTAGCTTATAGTTATGAATTTCTTCAAGTTAACAATTAAATCAATTCCTTTATCC
TTTTGTTGTCAATTGAAGCAATTCCCAGTGCGTTGTGGCTTTGTTTGTATGATTGACCTTTTTATCAACAATATTTG
AGCTTTCTTGTTGGACTACAAAACAAACTCCGAAAGGCCAAATGAACCAAATTGTTAAATTCAACTTAATTTGGG
ACATGAACATTTTCCAGTTTCCTTTATACTGTAAACTGTCAAGTCTCTTTTAATCTTGTTACCATTTTCTGAACTAT
ATATATAATTCCCAGTTAACATATGTTTGATTTATCTATATTAACATCTATTTGATGCCAAGTTGGTTTGGTGGCC
CAACCTTTTAAATAATTTGTTTATTAAATACAAACCTTTTAAATAATTTGTTAGAGCGTACAATGGACTTCCCTTA
TGTTAAAAATCTGTATAGTATCAAAGTTATTAGTAACATTGTGATTAGAATAGTGAAGTTATATCATTATGTTTTA
TAAAACTGTTTCTTCCGTTAGATATCTAACTCGTACGCTCAGTTTTATAAGCTGTTGACATTTAGATTTTTCATTCA
AAATACTTATTAGCACTCATTCTACACAGAGGATTAAAATGATAATTTTCTATCATTGATACCTCTTATACTGCAT
AATTATCAAACAAATAAACACAGAGGGTAAAAATGATTCATTGCAATGCTCCCTTCCAGTTTTTGAGGTAAAAAT
GATTCATTCTAATGCTAGTTCCTGAACGTAATCTACCCTCTTGAAAAACAGAGAAAAAGCTCAATTAGGATGCCA
TATCATTTAACACAATCAGCCAAACAACAAGCCTAAAACTATACTATGCTTGCTTGAAAATGAGAAGATGATGTA
TAGTCTTNTNATAGGTATTTAGGCTCACCTTTTCTCTGTGTTATTAGCTCTCAGTCTAGTATTTATCTTATTGGCAG
GCTAAACCCTGCAAGAGTCTGTGGGTAAGTGGCATCAGCAGTTCCACTACCAAGGAAGATCTAGAGGAAGAATT
TTCCAAATTTGGTAAATTTGAGGATTTCAAATTCCTAAGAGATAAGAATACAGCATACATTGATTATTCCAGACT
AGATGATGCATCAAAGGCTCTTAAAATGATGCATGGCAAGAAAAGAGGGGGATCTGTCATTCGTCGACTATC
TAAGATCTCAACCTAAAAGAGTAAGTCATTGGTTCAAATGAATATTGTTAAAAAAAATTGGAATGTAAGGGTAA
AATGGTTATTTAGTTTTGTTGAGATGGTTTATTCAATCTAGAAAAGAAAAGAAAAGAAAAGTCTTAATGTACACA
ACACATGATACAGGACTTAATGGGGGGATTGTTACATTTCCACGTGTAAGTAACAATCCCCACCCACCCTTTTAA
CAATATTTAATAAATCAACATGTATTGAATATTTTATCCTGTTTTGACTATGTAACTAACAGGACCAGGGTCCTGA
TTTCCGTGATGCAAAGGAAGGACAGCATTTCAGAAGTATGGGTCCACATGATTCACCATGGTTACCTCCAGATAG
TGTTAAAAATTACCCTGATGTTTCATATTATGGTCCCAAAAGGCAACAGGTTATAACCTAGTTCTAGTTGTTGTTC
AATTTATGCCTTTTTTGCGTTTCCACAGAAGTAATGATATGATAGAAAAACATATGGCATCATTTACCTGTTATTT
TTCTATAGTGTCATTATTTCTAAAAATACACATTAGGGCGGGGTTGTTTCACAAAATATTTTTGGAAGGAGTTGAA
TCTTTTAAACAAATTGTAGTTTGNGGTGTGTTTGGTCGGTTAGCAGAAAGTGAAATGGAAATCATTATTAAAAAC
TGAATAGTTATAATTTTTGCCCTTATAACCAAGTTTTTTATTTTTGTTTATAGCAATTATGGGGTATTTTGGTTTTTT
AAGAAGNGGGTATTTTTATTTTTATATTTTTCCAATAATCGGGTATTTTGTTATNTTTNTNAAAAAAAAAAAACTT
CCATGCGATCCTAGGAACTGTGGAATCCATCAAACATGGAATGTGAAGGAATCTGAATCTGATGGAATTGAATT
GTGTTCATTTGCAACTAAACATGAAATGGAATGGAATGGAATGTGGCACGAATTCTAATTTACAACATAATTTGT
GGTGTTTTCACAGCATAATAATATATTGGGCATGGAAGGTCGAAAAGGAGACGGAGAGCAACCAAGTAATGTTT
TGGTAATATCATATCCTCCCATCGTTCACATCGACGAACAAATGTTACACAACGCCATGATTTTATTTGGTGAGAT
TGATGACATCACAAGCTTTCCTTCAAGACATCATTATCTTGTTGAATTCAGAAGCGTAGAAGAGGCACAGCTGGC
TAAAGACGGACTCCAAGGCCGTTTGTTTAACGACCCCAGAATCTCAATCATGTTTGCAACCAACGATCAAACACC
AAACAAAGATATCCCTGGGTTCCACCCTCATGGAATCTTTAATGAACCACCACCTCCACAGTTGGATGCATATGG
TCACCCTGTATTAATAATGCCTAATAGGCCTTATGTAGGAGGAGGACCTTTTGTCCCTCCAGACACGTTCGATCCT

CCAGAGTTTCCTCCTATGAGTATGAATATGAATCCAATGGGTGTCGGTGTCGGTGTCCCTAACTGGAGAAGGTCT
TCCCCTACCCCTAATCCGTCTCCTGGTCCATGGGATGTTTTTGATGGAACCCATTTGCATAGGGAACCTAAAAGAC
TAAGAACCGAGGGTAATATACCTTTGAGAGATATGAACGATCATGATCCATATGCTGCTGCTGGTGGTGGTGGTA
GACATCCTCCTGGTTGTGACTATATTTGGCGAGGCGTTATTGCTAAAGGTGGAACTCCTGTTTGTCATGCTCGTTG
TGTCCCTATTAGAGACTGGATTGGATATGAGATGTAAGTGATTTTCCTTTATAAGATGAACCAATTAATATATTTG
TGTAAATCACATTTATGATTTTTGTGACCCTGCAGCCCAGAAGTGGTTAATTGCTCAGCCAGAACAGGATTAGAT
ATGCTAGCAAAGCACTACACAGACGCAATCGGTTTCGATATCGTTTTCTTCTTGCCCGATAGTGAAGAAGACTTT
GCTTCATATACAGAATTTGTACGTTACTTAGGTGACAGAAACCGGGCCGGGGTTGCTAAATTCGATGACGGAACT
ACCCTTTTCCTGGTCCCACCTTCGGATTTCCTTTCCAATGTTTTAAATGTTTCCGGCCCCGAACGTTTGTATGGAGT
TATCCTCAAGTTCCCCCAACCACCACCACCACCACCTCAATACATCGATAAGCAGCAGATCCCTCCTCAAAC
TCTTCAAAATGACTACAATTTAATCATGCCAGGTGGCGGGGAGAAGCTGCTCCAAATTGATTATAGCGGTGTCCC
ACTCTCACATGAAGACTTAAAGTCACCGGCTCCTCAACCGCCGCCACCACAACCACAACCACAACCGGTGTCATC
TACGCTCTCTTTAACACCTGAACTTATCGCAACTCTCGCTTCTTTGGCCAAAGGAAAGTTTAACGGTGNCCAACA
GCAACCTTCCACCGCCACGTCAGCATCAGCATCAACAGTGGGGCCCGGGAGACCATGGGAATATGAGCCTGAAC
CTAGTGGACATTTAATGCAACAAGCTCAGCCACAAGTTCCACCGGGATATGGTGGTCAAGATGTGAATTTCACCA
TGCCGCAAGGGCAATCTGGTCATTTTGCAGTTCCAAGCCAGCAGCAATACGTACCCAATTTCCCGCAGGCCGGTG
GGTATGGATTTGAACAGAAGCCGGAAGCTTTAAGTCAGGTTTATGGTGGTGGTAACATTTATCAACCACAAAGTA
TGGGAGGTGAACAAAAGACGGATGCATCCTTGGGAAGTCAGATTTATGGTGGTGGTGGTAATGTTTACCAACCA
CAAAATATGGGAGGAGGTGAGAACTCTGGTGGTGGTGGTGTGCAGCTGCCGGAGCAAATGCAGCAGCTTCAGTC
TGCACTATATGCAGCGAGTCAACAGCAACCGGTGGCTGATTTTGATGCCGATAAAAACGAGAGATATCAGTCAA
CACTGCAGTTTGCTACCAATCTCCTCCTCAAAATACATCAGCAGCAGCCGGGGCAAGGTAGTGGTGCTGGTGGTT
CTCTTCACTGA

APPENDIX VII

Table A4 - Experimental plans - Larissa

	2006		2007				2008				2009				2010	
	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer
M2/M3 scored for bolting																
M3 rescored for bolting																
LB M3 back-crossed to WT (BC1)																
BC1 seed back-crossed to WT (BC2)																
BC2 self seed scored for bolting (BC2_S1)																
Self seed scored for homozygous mutant lines (BC2_S2)																
Homozygous mutant lines confirmed																
Mutant lines scored in field trial																
Mutant lines scored at different temperatures																

Table A5 - Experimental plans - Saladin

		2007			2008				2009				2010	
		Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer
Initial Trial	M2 scored for bolting													
	M3 seed collected													
	M2 rescored for bolting in GH/HG													
	LB M3 back-crossed to WT (BC1)													
	BC1 plants grown for self-seed (RZ)													
	BC1 seed back-crossed to WT (BC2)													
	BC2 plants grown for self-seed (RZ)													
	BC2 self seed scored for bolting													
		2007			2008				2009				2010	
		Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer
Second Population	M1 scored for bolting													
	M2 seed collected													
	M2 rescored for bolting in GH/HG													
	LB M2 back-crossed to WT (BC1)													
	BC1 plants grown for self-seed (RZ)													
	BC1 seed back-crossed to WT (BC2)													
	LB M3 back-crossed to WT (BC1)													
	BC2 plants grown for self-seed (RZ)													
	BC2 self seed scored for bolting													

APPENDIX VIII

ANOVA TABLES

From Chapter 3

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *AtFT* and nontransformed controls (T1 generation).

Table A7: *ft-1* transgenic lines (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	4148.314	4148.314	661.53	<.001
Residual	22	137.957	6.271		
Total	47	4160.979			

Table A8: L*er* WT plants (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	214.862	214.862	168.17	<.001
Residual	19	24.275	1.278		
Total	48	236.204			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFT* and nontransformed controls. (T1 generation).

Table A9: *ft-1* transgenic lines (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	2628.004	2628.004	512.12	<.001
Residual	19	97.500	5.132		
Total	44	2596.311			

Table A10: L*er* WT plants (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	44.387	44.387	23.07	<.001
Residual	23	44.250	1.924		
Total	52	139.321			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFT* and nontransformed controls. (T2 generation in LD conditions).

Table A11: *ft-1* transgenic lines (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	5920.932	1184.186	353.86	<.001
Residual	254	850.011	3.346		
Total	344	4095.258			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFT* and nontransformed controls. (T2 generation in SD conditions).

Table A12: *ft-1* transgenic lines (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	15565.02	3113.00	69.90	<.001
Residual	105	4676.43	44.54		
Total	141	16913.18			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFT* and nontransformed controls. (T2 generation in LD conditions).

Table A13: L*er* WT plants (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	431.837	86.367	32.33	<.001
Residual	386	1031.303	2.672		
Total	504	1524.448			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFT* and nontransformed controls. (T2 generation in SD conditions).

Table A14: L*er* WT plants (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	3	2327.81	775.94	61.37	<.001
Residual	47	594.24	12.64		
Total	99	2361.79			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFLK* and nontransformed controls. (T1 generation)

Table A15: *flk-4* transgenic lines (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	801.480	801.480	281.67	<.001
Residual	10	28.455	2.845		
Total	52	375.698			

Table A16: Col-0 WT plants (leaves to flower)

Source of variation	d.f	s.s.	m.s.	v.r.	F pr.
Line	1	63.431	63.431	52.62	<.001
Residual	7	8.438	1.205		
Total	20	71.810			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFLK* and nontransformed controls. (T2 generation)

Table A17: Col-0 WT plants (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	4	202.108	50.527	36.16	<.001
Residual	146	204.006	1.397		
Total	205	347.364			

ANOVA tables showing the significance of the differences in flowering time between plants transformed to over-express *LsFLD* and nontransformed controls. (T1 generation)

Table A18: *fld-3* transgenic lines (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	123.410	123.410	50.03	<.001
Residual	9	22.200	2.467		
Total	32	204.970			

Table A19: Col-0 WT plants (leaves to flower)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	25.984	25.984	17.38	<.001
Residual	29	43.350	1.495		
Total	72	139.479			

From Chapter 4

Table A20: ANOVA tables showing the significance of the differences in bolting time for the Larissa EMS M2/M3 population

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	59	47410.42	803.57	9.95	<.001
Residual	524	42301.73	80.73		
Total	583	89712.16			

Table A21: ANOVA tables showing the significance of the differences in bolting time for the Larissa EMS M2/M3 population – repeat screen under a natural LD photoperiod

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	6	20555.02	3425.84	58.04	<.001
Residual	17	1003.36	59.02		
Total	31	10781.88			

Table A22: ANOVA tables showing the significance of the differences in bolting time for Larissa lines S2_BC2_2_D3 / S2_BC2_2_G2 / S2_BC2_2_C1 compared to WT – with no vernalisation treatment

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	3	1819.819	606.606	89.24	<.001
Residual	56	380.644	6.797		
Total	78	2446.608			

Table A23: ANOVA tables showing the significance of the differences in bolting time for Larissa lines S2_BC2_2_D3 / S2_BC2_2_G2 / S2_BC2_2_C1 compared to WT – with a vernalisation treatment

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	3	2054.291	684.764	255.04	<.001
Residual	55	147.670	2.685		
Total	77	2405.385			

Table A24: ANOVA tables showing the significance of the differences in bolting time for Larissa line S2_BC2_164_E1compared to WT – with no vernalisation treatment

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	3276.100	3276.100	1836.16	<.001
Residual	19	33.900	1.784		
Total	39	3377.500			

Table A25: ANOVA tables showing the significance of the differences in bolting time for Larissa line S2_BC2_164_E1compared to WT – with a vernalisation treatment

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	38.025	38.025	19.81	<.001
Residual	19	36.475	1.920		
Total	39	134.775			

Table A26: ANOVA tables showing the significance of the differences in bolting time for Larissa line S2_BC2_307_B4 compared to WT – with no vernalisation treatment

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	2496.400	2496.400	1295.95	<.001
Residual	19	36.600	1.926		
Total	39	2611.600			

Table A27: ANOVA tables showing the significance of the differences in bolting time for Larissa line S2_BC2_307_B4 compared to WT – with a vernalisation treatment

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	2496.400	2496.400	956.28	<.001
Residual	19	49.600	2.611		
Total	39	2611.600			

Table A28: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT in the Wellesbourne field trial I

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	3845.300	769.060	546.72	<.001
Residual	95	133.635	1.407		
Total	139	2612.971			

Table A29: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT in the Wellesbourne field trial II

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	2526.580	505.316	112.17	<.001
Residual	84	378.412	4.505		
Total	128	1922.140			

Table A30: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT in the Spalding field trial I

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	4183.080	836.616	244.95	<.001
Residual	82	280.070	3.415		
Total	126	2415.039			

Table A31: ANOVA tables showing the significance of the differences in bolting time comparing each of the lines in each of the three field trial sites

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Line ignoring Trial	5	5010.234	1002.047	245.13	< 0.001
Line eliminating Trial	5	4678.827	935.765	228.92	< 0.001
Trial ignoring Line	2	32124.031	16062.016	3929.30	< 0.001
Trial eliminating Line	2	31792.625	15896.313	3888.76	< 0.001
Line.Trial	10	726.151	72.615	17.76	< 0.001
Residual	378	1545.172	4.088		
Total	395	39074.182	98.922		

Table A32: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT at 18°C

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	10112.88	2022.58	70.34	<.001
Residual	15	431.29	28.75		
Total	23	10658.62			

Table A33: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT at 21°C

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	10280.708	2056.142	241.98	<.001
Residual	15	127.458	8.497		
Total	23	10426.958			

Table A34: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT at 23°C

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	6207.88	1241.58	69.16	<.001
Residual	15	269.29	17.95		
Total	23	6533.62			

Table A35: ANOVA tables showing the significance of the differences in bolting time for the LB Larissa lines compared to WT at 25°C

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	6709.83	1341.97	56.15	<.001
Residual	15	358.50	23.90		
Total	23	7093.83			

Table A36: ANOVA tables showing the significance of the differences in bolting time for the M2 Saladin EMS ‘test’ population in the field

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	106	70626.2	666.3	4.90	<.001
Residual	124	16859.8	136.0		
Total	233	52445.7			

Table A37: ANOVA tables showing the significance of the differences in bolting time for the M2 Saladin EMS 'test' population in the glasshouse

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	18	13953.0	775.2	4.51	<.001
Residual	28	4814.6	171.9		
Total	49	14379.8			

Table A38: ANOVA tables showing the significance of the differences in bolting time for the M2 Saladin EMS 'test' population in the field – repeat screen

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	98	31229.26	318.67	4.75	<.001
Residual	213	14290.42	67.09		
Total	314	40638.71			

Table A39: ANOVA tables showing the significance of the differences in bolting time for the M2 Saladin EMS 'test' population in the glasshouse – residual seed

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	21	18082.38	861.07	16.61	<.001
Residual	13	673.87	51.84		
Total	36	6995.30			

Table A40: ANOVA tables showing the significance of the differences in bolting time for the LB M2 Saladin EMS 'main' population in the field

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	14	15977.38	1141.24	13.72	<.001
Residual	21	1747.37	83.21		
Total	38	12883.90			

Table A41: ANOVA tables showing the significance of the differences in bolting time for the LB M2 Saladin EMS 'main' population in the glasshouse

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	368.64	73.73	1.33	0.319
Residual	11	607.54	55.23		
Total	19	1058.55			

Table A42: ANOVA tables showing the significance of the differences in bolting time for the LB M2 Saladin EMS 'main' population in the haygrove

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	867	126559.41	145.97	8.04	<.001
Residual	2401	43579.86	18.15		
Total	3271	161233.80			

Table A43: ANOVA tables showing the significance of the differences in bolting time for the LB M3 Saladin EMS 'main' population in the glasshouse (includes lines representing where one, two or three plants were LB in the M2 population)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	69	7762.69	112.50	3.21	<.001
Residual	131	4594.17	35.07		
Total	203	9993.16			

Table A44: ANOVA tables showing the significance of the differences in bolting time for the T1/T2 Jessy population in the glasshouse under non-vernalised conditions

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	8	1777.19	222.15	16.48	<.001
Residual	15	202.26	13.48		
Total	28	1526.83			

Table A45: ANOVA tables showing the significance of the differences in bolting time for the T1/T2 Jessy population in the glasshouse under vernalised conditions

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	8	926.307	115.788	17.43	<.001
Residual	28	186.011	6.643		
Total	41	440.286			

Table A46: ANOVA tables showing the significance of the differences in bolting time for the T2/T3 Jessy population in the glasshouse under non-vernalised conditions

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	5	269.977	53.995	6.30	<.001
Residual	34	291.629	8.577		
Total	46	605.617			

Table A47: ANOVA tables showing the significance of the differences in bolting time for the T2/T3 Jessy population in the glasshouse under vernalised conditions

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	6	16.672	2.779	1.88	0.116
Residual	31	45.848	1.479		
Total	44	69.200			

Table A48: ANOVA tables showing the significance of the differences in bolting time for the BC2_S2_T2/T3 Jessy population in the glasshouse under vernalised conditions

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	17	151.666	8.922	1.68	0.079
Residual	50	265.780	5.316		
Total	70	493.972			

From Chapter 5

Table A49: ANOVA tables showing the significance of the differences in bolting time for the M2/M3 Larissa line 199 and WT in the glasshouse

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	890.32	890.32	79.62	<.001
Residual	10	111.82	11.18		
Total	22	1085.30			

Table A50: ANOVA tables showing the significance of the differences in bolting time for the M2/M3 Larissa line 185 and WT in the glasshouse

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	2	3262.20	1631.10	57.62	<.001
Residual	9	254.78	28.31		
Total	22	3801.74			

Table A51: ANOVA tables showing the significance of the differences in bolting time for the M3 Larissa line 185 and WT in the glasshouse – rescreen 1 under natural LD photoperiod

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	1	101.24	101.24	8.35	0.063
Residual	3	36.38	12.13		
Total	12	147.08			

Table A52: ANOVA tables showing the significance of the differences in bolting time for the M3 Larissa line 185 and WT in the glasshouse – rescreen 2 under natural LD photoperiod

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Line	2	1134.97	567.49	56.48	<.001
Residual	37	371.73	10.05		
Total	58	1870.17			

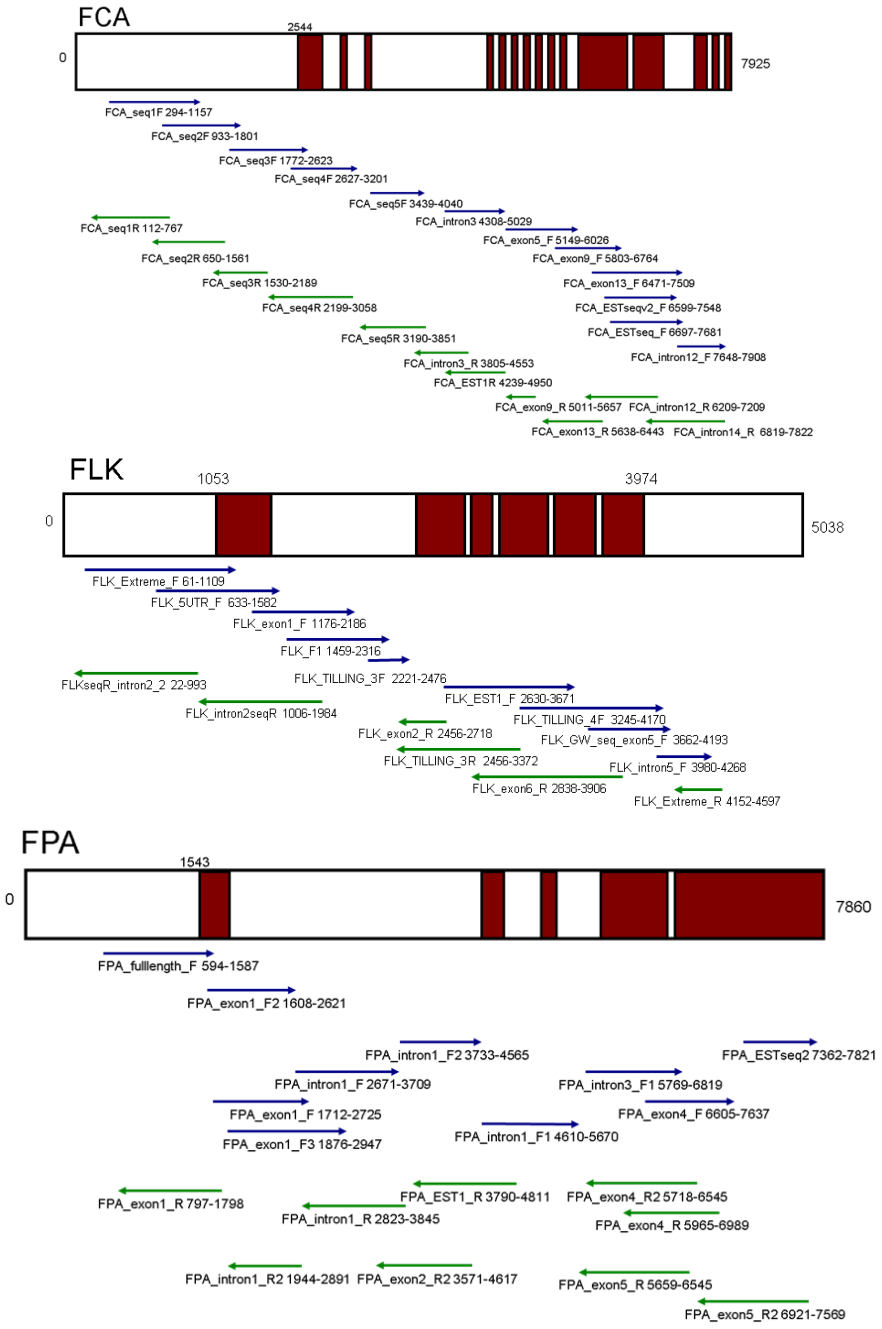
APPENDIX IX

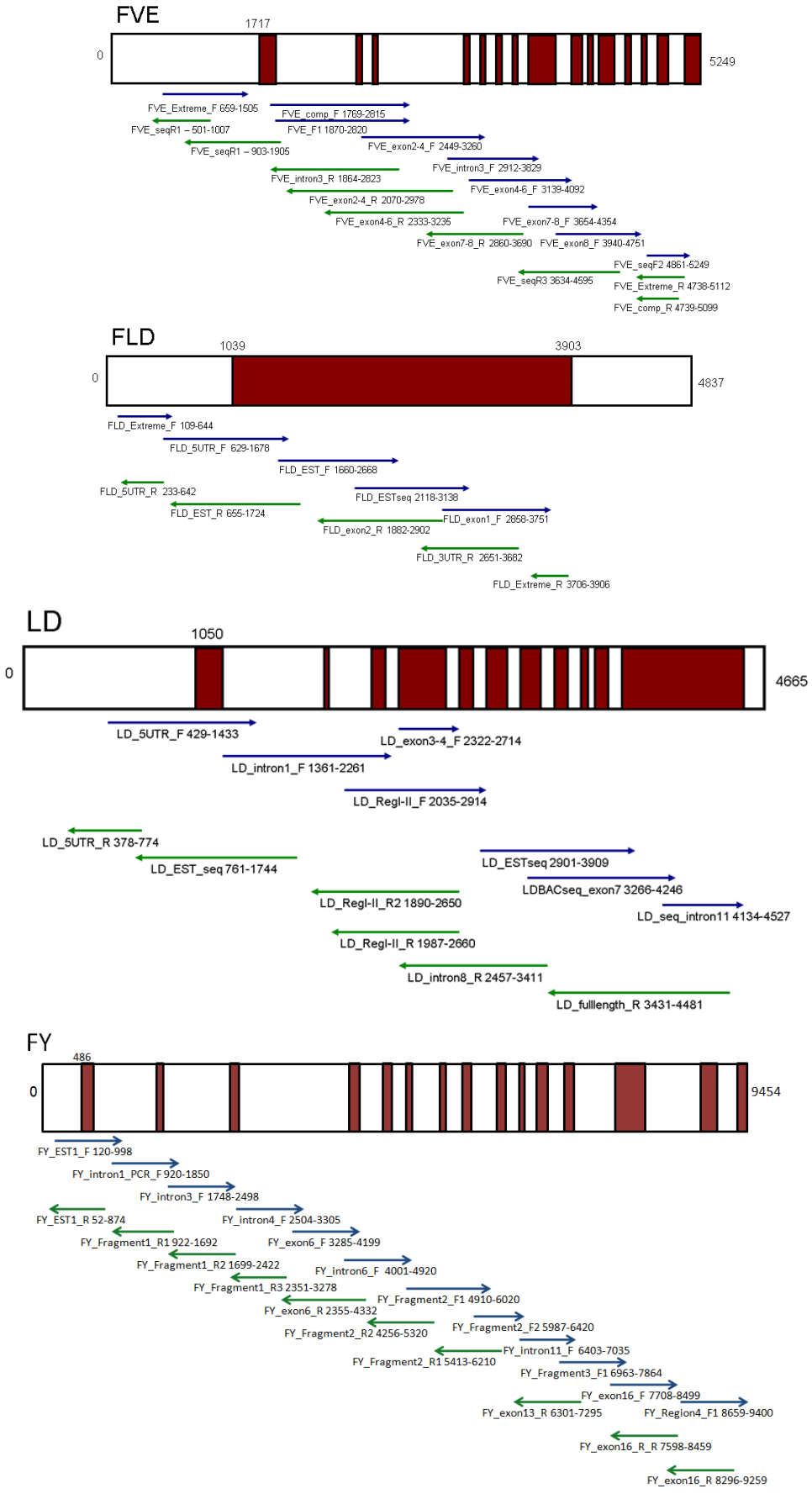
Table 54 - 96 well plate for TILLING experiments

	1	2	3	4	5	6	7	8	9	10	11	12
A	M3_CON4	M2_307_3	M2_320_4	M3_195_2	M3_126_6	M2_77_4	M3_44_11	M3_173_5	M3_87_4	M3_164_1	ser_ave_1567	M2_99_1
B	M2_69_8	M2_264_9	ser_LB_ 55	M2_354_3	M2_303_11	M3_363_10	M3_12_12	H2O	M2_2_10	M2_199_4	MARISKA_WT	M3_29_7
C	M2_90_9	M2_147_1	M3_37_5	M3_7_1	JESSY_WT	vir_LB_16201	M2_161_9	Mar_T1_53b	M2_307_3	M2_245_1	M2_5_3	vir_LB_1589
D	Jes_T2_Tnt1a_2	M3_164_1	M2_2_10	M3_143_9	M2_99_4	M3_CON5	vir_ave_W12	M3_164_1	M2_273_8	ser_ave_1897	M2_23_9	M2_2_10
E	M2_26_9	ef1ef2	M2_213_4	SAL _WT	M2_4_7	LAR_WT	M3_220_2	M3_105_2	M2_134_11	M3_71_4	M3_171_4	M3_193_6_1
F	M2_2_10	M2_352_5	M2_141_1	sag_LB_ 81	M2_61_11	M3_91_2	Mar_T1_53a_5	M3_265_9	sag_LB_36	M2_184_1	M2_207_5	M2_185_9
G	M2_307_3	M2_307_3	vir_LB_16198	M2_72_8	M3_63_1	M2_132_8	M2_130_10	M2_59_7	LAR_WT	M3_349_2	M3_164_1	vir_LB_16200
H	sag_ave_W92	M2_780_10	M2_310_12	H2O	M2_80_12	M3_117_6	vir_LB_16199	M2_110_1	M3_CON4	Jes_T2_Tnt1a_5	M2_335_9	H2O

APPENDIX XX

Figure A1 - Sequencing details of target genes




FVE
1717
0
5249
FVE_Extreme_F 659-1505
FVE_comp_F 1769-2815
FVE_seqR1 – 501-1007
FVE_F1 1870-2820
FVE_seqR1 – 903-1905
FVE_exon2-4_F 2449-3260
FVE_intron3_F 2912-3829
FVE_intron3_R 1864-2823
FVE_exon4-6_F 3139-4092
FVE_exon2-4_R 2070-2978
FVE_exon4-6_R 2333-3235
FVE_exon7-8_F 3654-4354
FVE_exon7-8_R 2860-3690
FVE_exon8_F 3940-4751
FVE_seqF2 4861-5249
FVE_seqR3 3634-4595
FVE_Extreme_R 4738-5112
FVE_comp_R 4739-5099
FLD
1039
3903
0
4837
FLD_Extreme_F 109-644
FLD_5UTR_F 629-1678
FLD_EST_F 1660-2668
FLD_5UTR_R 233-642
FLD_ESTseq 2118-3138
FLD_EST_R 655-1724
FLD_exon1_F 2858-3751
FLD_exon2_R 1862-2902
FLD_3UTR_R 2651-3682
FLD_Extreme_R 3706-3906
LD
1050
0
4665
LD_5UTR_F 429-1433
LD_exon3-4_F 2322-2714
LD_intron1_F 1361-2261
LD_RegI-II_F 2035-2914
LD_5UTR_R 378-774
LD_ESTseq 2901-3909
LD_EST_seq 761-1744
LDBACseq_exon7 3266-4246
LD_RegI-II_R2 1890-2650
LD_seq_intron11 4134-4527
LD_RegI-II_R 1987-2660
LD_intron8_R 2457-3411
LD_fulllength_R 3431-4481
FY
486
0
9454
FY_EST1_F 120-998
FY_intron1_PCR_F 920-1850
FY_intron3_F 1748-2498
FY_EST1_R 52-874
FY_intron4_F 2504-3305
FY_Fragment1_R1 922-1692
FY_exon6_F 3285-4199
FY_Fragment1_R2 1699-2422
FY_intron6_F 4001-4920
FY_Fragment1_R3 2351-3278
FY_Fragment2_F1 4910-6020
FY_exon6_R 2355-4332
FY_Fragment2_R2 4256-5320
FY_Fragment2_F2 5987-6420
FY_intron11_F 6403-7035
FY_Fragment2_R1 5413-6210
FY_Fragment3_F1 6963-7864
FY_exon16_F 7708-8499
FY_exon13_R 6301-7295
FY_Region4_F1 8659-9400
FY_exon16_R_R 7598-8459
FY_exon16_R 8296-9259